

From: Leffers, Gerald  
Sent: Tuesday, January 25, 2005 4:06 PM  
To: STIC-Biotech/ChemLib  
Subject: FW:10/776,213 sequence search

Examiner's mailbox is in 2c70 of Remsen. Thank you. Gerry Leffers

*Gerald G. Leffers Jr., PhD*  
Primary Examiner, Art Unit 1636  
Remsen Building, Room 02A69  
(571) 272-0772

-----Original Message-----

From: Leffers, Gerald  
Sent: Tuesday, January 25, 2005 4:02 PM  
To: STIC-Biotech/ChemLib  
Subject: 09/776,213 sequence search

Please do a search/interference search for SEQ ID NO: 2 of this application (~723 nucleotides). Claims are to sequences comprising as few as 17 consecutive nucleotides of SEQ ID NO: 2. Thank you. Gerry Leffers

*Gerald G. Leffers Jr., PhD*  
Primary Examiner, Art Unit 1636  
Remsen Building, Room 02A69  
(571) 272-0772

RECEIVED  
JAN 25 2005  
STIC-Biotech/ChemLib  
(STIC)

\*\*\*\*\*

STAFF USE ONLY

Searcher: \_\_\_\_\_  
Searcher Phone: 2- \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Rev. Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search

NA Sequence: # \_\_\_\_\_  
AA Sequence: # \_\_\_\_\_  
Structure: # \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: February 7, 2005, 21:10:02 ; Search time 3600 Seconds

(without alignments)  
9731.423 Million cell updates/sec

Title: US-10-776-213-2

Perfect score: 723  
Sequence: 1 ctttcgattagcagcagcagc.....aaagacataataacat 723

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenBank: \*  
1: gb\_ba: \*  
2: gb\_hg: \*  
3: gb\_in: \*  
4: gb\_cm: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sbs: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	723	100.0	723	6	AR492027 Sequence
2	723	100.0	11427	6	AR492049 Sequence
3	723	100.0	13073	6	AR492045 Sequence
4	720.4	99.6	850	6	AR492055 Sequence
5	720.4	99.6	23498	8	SC9920 Sequence
6	497.4	68.8	680	6	AX536716 Sequence
7	47	6.5	7218	6	166494 Sequence
8	46.8	6.5	257757	3	AE014837 Sequence
9	43.6	6.1	170295	5	BE663508 Sequence
10	43.6	6.0	171265	5	AC092573 Homo sapi
11	42.8	5.9	208729	2	AC012052 Homo sapi
12	42.8	5.9	80659	9	AC004553 Homo sapi
13	42.8	5.9	195932	2	AC021710 Homo sapi
14	42	5.8	28858	9	AL359542 Human DNA
15	41.8	5.8	209764	10	AL645468 Human DNA
16	41.2	5.7	3291	8	AF210049 Petunia x
17	40.6	5.6	183978	2	AC024967 Homo sapi
18	40.6	5.6	36763	2	AC149338 Plakophilin
19	40.4	5.6	248333	2	AC097969 Rattus no

20	40.4	5.6	261990	2	AC111513 Rattus no
21	40	5.5	90500	9	AL157889 Human DNA
22	40	5.5	139604	9	AC090639 Homo sapi
23	40	5.5	208083	2	AC009649 Homo sapi
24	40	5.5	227184	2	AC009868 Homo sapi
25	40	5.5	238936	2	AC111758 Rattus no
26	39.8	5.5	253105	3	PFMA1397 Plasmid
27	39.4	5.4	152886	2	AC110665 Canis fam
28	39.4	5.4	157223	9	AC099664 Homo sapi
29	39.4	5.4	225885	2	AC113236 Canis fam
30	39	5.4	2000	6	AX655393 Sequence
31	39	5.4	110000	8	CR382131_37 Continuation (38 o
32	39	5.4	150244	9	AC091952 Homo sapi
33	39	5.4	168186	2	AC146425 Pen trogl
34	39	5.4	174007	2	AC136942 Homo sapi
35	38.8	5.4	110000	2	AC137108 Rattus no
36	38.8	5.4	129547	8	AC006434 Genomic s
37	38.8	5.4	137625	9	AC104663 Homo sapi
38	38.8	5.4	164310	2	AC016156 Homo sapi
39	38.8	5.4	181851	10	AC102413 Mus muscu
40	38.8	5.4	203230	2	AC026346 Homo sapi
41	38.8	5.4	205592	2	AC150295 Callithrix
42	38.8	5.4	229634	2	AC150010 Callithrix
43	38.8	5.4	239450	2	AC106212 Rattus no
44	38.6	5.3	2311	6	AX664151 Sequence
45	38.6	5.3	100571	9	AL161797 Human DNA

#### ALIGNMENTS

RESULT 1  
LOCUS AR492027 723 bp DNA linear PAT 15-MAY-2004  
DEFINITION Sequence 2 from patent US 6716601.  
ACCESSION AR492027  
VERSION AR492027.1 GI:47260496  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 723)  
AUTHORS Belfield G.P. and Oakley, C.  
TITLES Compositions and methods utilizing the yeast ZEO1 promoter  
JOURNAL Patent: US 6716601-A 2 06-APR-2004;  
FEATURES  
source Location/Qualifiers  
1..723  
/organism="unknown"  
/mol\_type="genomic DNA"

#### ORIGIN

Query Match 100.0%; Score 723; DB 6; Length 723;  
Best Local Similarity 100.0%; Pred. No. 3.8e-207;  
Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CTTTGATTAGCAGCAGCAGCATCATGACTGCGTATATAAATACATACGAAAA	60
DB	1	CTTTGATTAGCAGCAGCAGCATCATGACTGCGTATATAAATACATACGAAAA	60
QY	61	ACCATTAAGAGCAAGCGATCTCTTGGAGGAAAGAGAGAGCGCTGTAAAGGGGAT	120
DB	61	ACCATTAAGAGCAAGCGATCTCTTGGAGGAAAGAGAGAGCGCTGTAAAGGGGAT	120
QY	121	GGGGGCTAAGAGTCAATTCATCTTCTTCCCTGCGGTCGAGCCCGGAGCCCTCTCT	180
DB	121	GGGGGCTAAGAGTCAATTCATCTTCTTCCCTGCGGTCGAGCCCGGAGCCCTCTCT	180
QY	181	CTCCCGGCAAGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	240
DB	181	CTCCCGGCAAGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	240
QY	241	GCACATGACATTAATGCTGCTGAGCATCTGAGTGGCTGAGTGGTGTATCTCACT	300
DB	241	GCACATGACATTAATGCTGCTGAGCATCTGAGTGGCTGAGTGGTGTATCTCACT	300

Examined Search Notes  
10/776,213

Db 241 GCACATGACTAATATGCTGAGACATCTCCATGAGCTGTGTGTGTAATCTCACAGT 300  
Qy 301 GGTAAACGGCACCGTGGCTCGGAAACGTTTCTTGTGCAAAATTTAGAAACAGGGGCTACA 360  
Db 301 GGTAAACGGCACCGTGGCTCGGAAACGTTTCTTGTGCAAAATTTAGAAACAGGGGCTACA 360  
Qy 361 GTCCTGATATAGAAATTAAGCGCATTTTGTGCTAGAGCGCGCCCGGGGCGCCCTTCCCA 420  
Db 361 GTCCTGATATAGAAATTAAGCGCATTTTGTGCTAGAGCGCGCCCGGGGCGCCCTTCCCA 420  
Qy 421 ATAGGAGGCGCAGTTTATCGCGGAGCTCTACTTCTTCTTATTTGGTAAAGCCCTTTC 480  
Db 421 ATAGGAGGCGCAGTTTATCGCGGAGCTCTACTTCTTCTTATTTGGTAAAGCCCTTTC 480  
Qy 481 TGTTTTCGGCCAGTGTGCTGCAAGCTGTGCGCGCCGAGAACATAGTGAATAGGAGTGTAC 540  
Db 481 TGTTTTCGGCCAGTGTGCTGCAAGCTGTGCGCGCCGAGAACATAGTGAATAGGAGTGTAC 540  
Qy 541 TTTTCGATGAGAAATTAAGCAAGCGGAAAAAACTATGCGCTAGCGGAGTTGTTTTCAA 600  
Db 541 TTTTCGATGAGAAATTAAGCAAGCGGAAAAAACTATGCGCTAGCGGAGTTGTTTTCAA 600  
Qy 601 TCATATTAAGGAGAAATTTGCTCTCACTATGTGACAGTTTCTGGGACGTCTTAACCTTT 660  
Db 601 TCATATTAAGGAGAAATTTGCTCTCACTATGTGACAGTTTCTGGGACGTCTTAACCTTT 660  
Qy 661 TATTGCAAGAGCATTAATCATACATATTTCTCAAAAAAAGACTAATATTA 720  
Db 661 TATTGCAAGAGCATTAATCATACATATTTCTCAAAAAAAGACTAATATTA 720  
Qy 721 CAT 723  
Db 721 CAT 723

RESULT 2  
AR492049 11427 bp DNA linear PAT 15-MAY-2004  
LOCUS AR492049  
DEFINITION Sequence 24 from patent US 6716601.  
ACCESSION AR492049  
VERSION AR492049.1 GI:47260518  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 11427)  
AUTHORS Belfield,G.P. and Oakley,C.  
TITLE Compositions and methods utilizing the yeast ZEO1 promoter  
JOURNAL Patent: US 6716601-A 24 06-APR-2004;  
FEATURES  
Location/Qualifiers  
1..11427  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 100.0%; Score 723; DB 6; Length 11427;  
Best Local Similarity 100.0%; Pred. No. 6.1e-207;  
Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTCGATTAGCAGCAGACATCATAGACTGGCTATTAATAATACATACGAAAA 60  
Db 15 CTTTCGATTAGCAGCAGCAGACATCATAGACTGGCTATTAATAATACATACGAAAA 74  
Qy 61 ACCATAAAGACAAAGCAGTACTTCTGGAAGAAAGAGAGCAGCTGTGAAGGGAGAT 120  
Db 75 ACCATAAAGACAAAGCAGTACTTCTGGAAGAAAGAGAGCAGCTGTGAAGGGAGAT 134  
Qy 121 GGGGGCTAAGAGTCACTTCTTTTCCCTCGCGGTCCGAGCCGGGAGCCCTCT 180  
Db 135 GGGGGCTAAGAGTCACTTCTTTTCCCTCGCGGTCCGAGCCGGGAGCCCTCT 194  
Qy 181 CTCGCCGACAGATTTCTCTTCAATATCTCTTTATTTCCATCCGTTGAAGGAAC 240

Db 195 CTCGCCGACAGATTTCTCTTCAATATCTCTTTATTTCTATCCCGTTGAAGAAC 254  
Qy 241 GCACATGACTAATATGCTGAGACATCTCCATGAGCTGTGTGTGTAATCTCACAGT 300  
Db 255 GCACATGACTAATATGCTGAGACATCTCCATGAGCTGTGTGTGTAATCTCACAGT 314  
Qy 301 GGTAAACGGCACCGTGGCTCGGAAACGTTTCTTGTGCAAAATTTAGAAACAGGGGCTACA 360  
Db 315 GGTAAACGGCACCGTGGCTCGGAAACGTTTCTTGTGCAAAATTTAGAAACAGGGGCTACA 374  
Qy 361 GTCCTGATATAGAAATTAAGCGCAATTTTGTACAGCGCGCCCGGGGCGCCCTTCCCA 420  
Db 375 GTCCTGATATAGAAATTAAGCGCAATTTTGTACAGCGCGCCCGGGGCGCCCTTCCCA 434  
Qy 421 ATAGGAGGCGCAGTTTATCGCGGAGCTCTACTTCTTCTTATTTGGTAAAGCCCTTTC 480  
Db 435 ATAGGAGGCGCAGTTTATCGCGGAGCTCTACTTCTTCTTATTTGGTAAAGCCCTTTC 494  
Qy 481 TGTTTTCGGCCAGTGTGCTGCAAGCTGTGCGCGCCGAGAACATAGTGAATAGGAGTGTAC 540  
Db 495 TGTTTTCGGCCAGTGTGCTGCAAGCTGTGCGCGCCGAGAACATAGTGAATAGGAGTGTAC 554  
Qy 541 TTTTCGATGAGAAATTAAGCAAGCGGAAAAAACTATGCGCTAGCGGAGTTGTTTTCAA 600  
Db 555 TTTTCGATGAGAAATTAAGCAAGCGGAAAAAACTATGCGCTAGCGGAGTTGTTTTCAA 614  
Qy 601 TCATATTAAGGAGAAATTTGCTCTCACTATGTGACAGTTTCTGGGACGTCTTAACCTTT 660  
Db 615 TCATATTAAGGAGAAATTTGCTCTCACTATGTGACAGTTTCTGGGACGTCTTAACCTTT 674  
Qy 661 TATTGCAAGAGCATTAATCATACATATTTCTCAAAAAAAGACTAATATTA 720  
Db 675 TATTGCAAGAGCATTAATCATACATATTTCTCAAAAAAAGACTAATATTA 734  
Qy 721 CAT 723  
Db 735 CAT 737

RESULT 3  
AR492045 13073 bp DNA linear PAT 15-MAY-2004  
LOCUS AR492045  
DEFINITION Sequence 20 from patent US 6716601.  
ACCESSION AR492045  
VERSION AR492045.1 GI:47260514  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 13073)  
AUTHORS Belfield,G.P. and Oakley,C.  
TITLE Compositions and methods utilizing the yeast ZEO1 promoter  
JOURNAL Patent: US 6716601-A 20 06-APR-2004;  
FEATURES  
Location/Qualifiers  
1..13073  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 100.0%; Score 723; DB 6; Length 13073;  
Best Local Similarity 100.0%; Pred. No. 6.3e-207;  
Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTCGATTAGCAGCAGCAGACATCATAGACTGGCTATTAATAATACATACGAAAA 60  
Db 16 CTTTCGATTAGCAGCAGCAGCAGACATCATAGACTGGCTATTAATAATACATACGAAAA 75  
Qy 61 ACCATAAAGACAAAGCAGTACTTCTGGAAGAAAGAGAGCAGCTGTGAAGGGAGAT 120  
Db 76 ACCATAAAGACAAAGCAGTACTTCTGGAAGAAAGAGAGCAGCTGTGAAGGGAGAT 135  
Qy 121 GGGGGCTAAGAGTCACTTCTTTTCCCTCGCGGTCCGAGCCGGGAGCCCTCT 180

```

Db      136 GGGGGGTTAAGAAATCACTTCTTTTCCCTCGCGGTCGGACCCGGGACCCCTCCT 195
Qy      181 CTCCTCCGCAAGATTTCTTCTTTCATATCTCTTTTATTCCTATCCCGTTGAAGAAC 240
Db      196 CTCCTCCGCAAGATTTCTTCTTTCATATCTCTTTTATTCCTATCCCGTTGAAGAAC 255
Qy      241 GCACTATGACTAATATGCTGTGACATCTTCATAGCTGTGACTTTGTGTATCTCAAGT 300
Db      256 GCACTATGACTAATATGCTGTGACATCTTCATAGCTGTGACTTTGTGTATCTCAAGT 315
Qy      301 GGTAAACGGACCGCTGCTCGGAAACGGTTCCTTGTGACAACTCTAAGACAGGGCTTACA 360
Db      316 GGTAAACGGACCGCTGCTCGGAAACGGTTCCTTGTGACAACTCTAAGACAGGGCTTACA 375
Qy      361 GTCTCGATATATGAAATATTAAGGCAATTTTGTCTAGCGCGCGCGCGCCGCTTTCCCA 420
Db      376 GTCTCGATATATGAAATATTAAGGCAATTTTGTCTAGCGCGCGCGCGCCGCTTTCCCA 435
Qy      421 ATAGGAGGCGCAGTTTATCGCGGAGCTCTACTTCTTCTATTTGGGTAAAGCCCTTTC 480
Db      436 ATAGGAGGCGCAGTTTATCGCGGAGCTCTACTTCTTCTATTTGGGTAAAGCCCTTTC 495
Qy      481 TGTTTTCGCGCAGTGTGCTGTGACAGCTGCGCGGAGCAATAGTATTAAGGATTAAC 540
Db      496 TGTTTTCGCGCAGTGTGCTGTGACAGCTGCGCGGAGCAATAGTATTAAGGATTAAC 555
Qy      541 TTTTCATGAGAGATTTAGCAAGCGGAAAAAACTATAGCTAGCTGGAGTTGTTTTTCAA 600
Db      556 TTTTCATGAGAGATTTAGCAAGCGGAAAAAACTATAGCTAGCTGGAGTTGTTTTTCAA 615
Qy      601 TCATATTAAGAGAGAAATTTGTTGCTCATATGTGACAGTTTCTGGAGCTCTTAACCTT 660
Db      616 TCATATTAAGAGAGAAATTTGTTGCTCATATGTGACAGTTTCTGGAGCTCTTAACCTT 675
Qy      661 TATTGACAGAGACTATCAATCATACAGATATGTCAAAAAAAGAGCTAATATATA 720
Db      676 TATTGACAGAGACTATCAATCATACAGATATGTCAAAAAAAGAGCTAATATATA 735
Qy      721 CAT 723
Db      736 CAT 738

RESULT 4
AR492055      850 bp      DNA      linear      PAT 15-MAY-2004
LOCUS         AR492055      Sequence 30 from patent US 6716601.
DEFINITION    AR492055
ACCESSION     AR492055
VERSION       AR492055.1 GI:47260524
KEYWORDS
SOURCE        Unknown.
ORGANISM      Unknown.
REFERENCE     1 (bases 1 to 850)
AUTHORS       Belfield,G.P. and Oakley,C.
TITLES        Compositions and methods utilising the yeast ZEO1 promoter
JOURNAL       Patent: US 6716601-A 30 06-APR-2004;
FEATURES      Location/Qualifiers
               1..850
               /organism="Unknown"
               /mol_type="genomic DNA"

ORIGIN
Query Match      99.6%; Score 720.4; DB 6; Length 850;
Best Local Similarity 99.9%; Pred. No. 2,4e-206;
Matches 721; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 CTTTCGATTAGCAGCAGACATCATCATAGCTGCTCATAAAAATACACTACGAAAA 60
Db      78 CTTTCGATTAGCAGCAGACATCATCATAGCTGCTCATAAAAATACACTACGAAAA 137
Qy      61 ACCATAAAGAGCAAGGATACCTACTGTAAGAGAAAGAGACGCTTGTAAGGGGGAT 120

```

```

Db      138 ACCATAAAGAGCAAGGATACCTACTTGGAAGAAAGAGACAGCTTGTAAGGGGAT 197
Qy      121 GGGGGGTTAAGAAATCACTTCTTTTCCCTTCCGCGGTCCGACCCGGGACCCCTCCT 180
Db      198 GGGGGGTTAAGAAATCACTTCTTTTCCCTTCCGCGGTCCGACCCGGGACCCCTCCT 257
Qy      181 CTCCTCCGCAAGATTTCTTCTTTCATATCTCTTTTATTCCTATCCCGTTGAAGAAC 240
Db      256 CTCCTCCGCAAGATTTCTTCTTTCATATCTCTTTTATTCCTATCCCGTTGAAGAAC 317
Qy      241 GCACTATGACTAATATGCTGTGACATCTTCATAGCTGTGACTTTGTGTATCTCAAGT 300
Db      318 GCACTATGACTAATATGCTGTGACATCTTCATAGCTGTGACTTTGTGTATCTCAAGT 377
Qy      301 GGTAAACGGACCGCTGCTCGGAAACGGTTCCTTGTGACAACTCTAAGACAGGGCTTACA 360
Db      378 GGTAAACGGACCGCTGCTCGGAAACGGTTCCTTGTGACAACTCTAAGACAGGGCTTACA 437
Qy      361 GTCTCGATATATGAAATATTAAGGCAATTTTGTCTAGCGCGCGCGCGCCGCTTTCCCA 420
Db      438 GTCTCGATATATGAAATATTAAGGCAATTTTGTCTAGCGCGCGCGCGCCGCTTTCCCA 497
Qy      421 ATAGGAGGCGCAGTTTATCGCGGAGCTCTACTTCTTCTATTTGGGTAAAGCCCTTTC 480
Db      498 ATAGGAGGCGCAGTTTATCGCGGAGCTCTACTTCTTCTATTTGGGTAAAGCCCTTTC 557
Qy      481 TGTTTTCGCGCAGTGTGCTGTGACAGCTGCGCGGAGCAATAGTATTAAGGATTAAC 540
Db      558 TGTTTTCGCGCAGTGTGCTGTGACAGCTGCGCGGAGCAATAGTATTAAGGATTAAC 617
Qy      541 TTTTCATGAGAGATTTAGCAAGCGGAAAAAACTATAGCTAGCTGGAGTTGTTTTTCAA 600
Db      618 TTTTCATGAGAGATTTAGCAAGCGGAAAAAACTATAGCTAGCTGGAGTTGTTTTTCAA 677
Qy      601 TCATATTAAGAGAGAAATTTGTTGCTCATATGTGACAGTTTCTGGAGCTCTTAACCTT 660
Db      678 TCATATTAAGAGAGAAATTTGTTGCTCATATGTGACAGTTTCTGGAGCTCTTAACCTT 737
Qy      661 TATTGACAGAGACTATCAATCATACAGATATGTCAAAAAAAGAGCTAATATATA 720
Db      738 TATTGACAGAGACTATCAATCATACAGATATGTCAAAAAAAGAGCTAATATATA 797
Qy      721 CA 722
Db      798 AA 799

RESULT 5
SC9920
LOCUS         SC9920      23498 bp      DNA      linear      PLN 11-AUG-1997
DEFINITION    S.cerevisiae chromosome XIII cosmid 9920.
ACCESSION     Z48639.1 GI:732924
VERSION       Z48639.1 GI:732924
KEYWORDS      COX7: cytochrome oxidase; delta element; glutamate decarboxylase;
SOURCE        Saccharomyces cerevisiae (baker's yeast)
ORGANISM      Saccharomyces cerevisiae
REFERENCE     1 (bases 1 to 23498)
AUTHORS       Hunt,S. and Bowman,S.
JOURNAL       Unpublished
               2 (bases 1 to 23498)
REFERENCE     Barrett,B., Rajandream,M.A. and Walsh,S.V.
AUTHORS       Direct Submission
TITLES        Submitted (10-MAR-1995) Saccharomyces cerevisiae chromosome XIII
JOURNAL       sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
               CB10 1RO E-mail: barrell@sanger.ac.uk
Notes:
All CDS over 100 codons have been analysed. CDS that are completely
overlapped and those that are overlapped by more than 50%
of their length by a larger CDS have been omitted from this
analysis.

```

Details of the omitted CDS are available on request. The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons and the calculated codon adaptation index (CAI) is given for each CDS. Cosmid 9920 is overlapped at the start of this sequence by cosmid 9408 and at the end of this sequence by cosmid 8156.

FEATURES  
source location/Qualifiers  
1..23498  
/organism="Saccharomyces cerevisiae"  
/mol\_type="genomic DNA"  
/strain="AB972"  
/db\_xref="taxon:4932"  
/chromosome="XIII"  
/map="13R"  
/clone="cosmid 9920"  
complement (<1..2870)  
/note="YM9920.01c, unknown, partial, len: 956, CAI: 0.14, PS00061 Short-chain alcohol dehydrogenase family signature"

CDS  
/product="unknown"  
/protein\_id="CA88574.1"  
/db\_xref="GI:732925"  
/translation="MSFGINTFOOYNTDGLGNGVRLSNYPDGLPDPGLNSLYS  
NELKLPKSLRDETTKEKALMDLSNLSDPNORBYFNDFILCWSQIYAKLIID  
KPAKILQSHQITIMLYKSLRKIKSKIDPFLILGTCLEIVSKSRKSLNTECN  
LKLPLVHNDVSEBSSSLKVLISDSIWKLLNKGONTAYETVALIDVLTREG  
MSKYLKMLAVKAKLLKSLVHTSNTKILKQVPLPSILNLTATLDYDEDTGYS  
SEKYLKPLASVRSIPSPGFNAVALYSYTKRSFLDYLEM.PPMOKVQVLNKG  
BARKSAEVLNFWTNPLKFADESEKVKQWSEI FNLSCGKLSSETYKALDLS  
GVFPDPKWERELIEDYDSDIIRIKIVFEKPLPALLTPNNSAISRLPFDVQIL  
ETDPNSVNRKYDGVYDALYFLDSMIFLNGKIKGTINEPILVQESTYQNFAGIQA  
LYNSKFPKNTDITSLDFPVALSPMLPTTITLATNBLNDIYQOLKMSDLEIE  
FSTLLNTDPLSCALYSEEDTNEKLPRLSLQAKNSEIANKLAQVILQHAQVYSP  
GAKKYVTHAVELINGCNDTSQIFFPNALIEVPARYPAIDYSSLSVLSNTHTLS  
DLYDGHFTFRHKVNLNFSDIIVGNYIOPANGDAMLTPDIASNSVYFFIYSRVLY  
KVLNSIDTVSTTNGILASVESFVTKIVDQKSDKQYL"

misc\_feature  
complement (2406..2438)  
/note="PS00061 Short-chain alcohol dehydrogenase family signature"

tRNA

LTR  
complement (3200..3272)  
/product="tRNA-Ala"  
/note="tRNA-Ala, anticodon agc, len: 73"

CDS

complement (3379..3706)  
/note="delta element"  
complement (join(4101..4196,4691..4780))  
/note="YM9920.02c, unknown, len: 61, CAI: 0.17, possible small spliced gene"

CDS

complement (start=1)  
/product="unknown"  
/protein\_id="CA88575.1"  
/db\_xref="GI:732926"  
/translation="MGELAPMKNKPDGITTNNQSLYRQIRMSTIGLRLLPLNNKAI  
IPNLALALHLCALYTE"  
complement (join(4101..4196,4434..4505))  
/note="YM9920.03c, unknown, len: 55, CAI: 0.13, possible small spliced gene"  
/codon\_start=1  
/product="unknown"  
/protein\_id="CA88576.1"  
/db\_xref="GI:732927"  
/translation="UniProt/TREMBL:O8X1Y9"  
/db\_xref="UniProt/TREMBL:O8X1Y9"  
ALAHCAIYTE"  
complement (4197..4199)  
/note="possible splice acceptor sequence, cag"

misc\_feature  
complement (4222..4228)  
/note="possible splice branch sequence, tacaac"  
complement (4428..4433)  
/note="possible donor sequence, gtagtc"  
complement (4685..4690)  
/note="possible donor sequence, gtacgc"  
complement (4699..4694)  
/note="possible donor sequence, gtacgc"  
5632..7389  
/note="YM9920.04, unknown, len: 585, CAI: 0.18, putative glutamate decarboxylase gene"

CDS

complement (start=1)  
/product="unknown"  
/protein\_id="CA88577.1"  
/db\_xref="GI:732928"  
/db\_xref="GOA:004792"  
/db\_xref="SGD:50004862"  
/translation="MLHRRGSKQKNPENIAGKVHDLAQLQLSNDVQSAVQSHOC  
SNMRDTSQGMANKYVPRKGLPADLILHNEILTDGNPHNLASFNVTPTTDDA  
RKLIDENLTKNLADNDEYDQILTELTORCISMLQMHANPDEBPIGATTSSTAIML  
GGIAMKKEWHEMKNAKDASKNTIMSSACOVALLKRTTFEYECRLVPVSRSHM  
LPESILMDYDENTIGCFVILGTYTGHLNVEKYADLSQIEKADHDMASNTDIPHA  
DASGGFTIPGFEKEHMKAYGMRNCGNHRVNSMTSGHKFGLTPPGVLMWRE  
SLADELRPKLYLGVEYETGMLNPSRPFQVNVQYENFVSLGSGYRTOFQNSLFWA  
RASPRLNASSCLPGCFEIVSSIHESIENSAFQVNDMEHPQAYRGVPLVAFKLS  
KTFHEYBEVQALISLNRGWIIPNYLPKATDSDSDEKVELRVVFRSEMLDLAQ  
LIVDESILTKLHSYEKCHHIELASEQTPERKSSFIEMLALASPDODLPTPDE  
LEKNKLEIETTRNYRGIC"  
7746..8846  
/note="YM9920.05, unknown, len: 366, CAI: 0.11, similar to YK56 YERST P36156, YKR076W, hypothetical protein in SIS2, 62.6% identity in 369 aa overlap"

CDS

complement (start=1)  
/product="unknown"  
/protein\_id="CA88578.1"  
/db\_xref="GI:732929"  
/db\_xref="SGD:50004863"  
/translation="MSRKSANRKAFAKQSSPFRITISADHPYKPAKRYMLYVAL  
PCPMAQRTTLPRLKGLAPIIGCSVAHMLDDKGRFLDEGQKTNBHMWFIAGIS  
SVNLNTPSTPARNNAHKLVDGTEPHYGRSLDFEFTKPDYKGRFTVPVMDL  
EETCIYNNBSDDIIGLMSAARDEFGEYRQVRLVPSLEAKITDFNSWYVDKING  
VYTAGFAECAEYEREVTSLPQYLDKLEMLDKTDLDEAEYKNNKXKITDRYALIG  
DTLEADVRLYPTIVRFDVVYHQHFKCNATIRDDYSNHHWMLKNIYRHRNAPORTTD  
FTHIKLGYTRSGPRVPIGITPLGPDIRPP"  
complement (10146..10550)  
/note="YM9920.06c, unknown, len: 134, CAI: 0.13"  
/codon\_start=1  
/product="unknown"  
/protein\_id="CA88579.1"  
/db\_xref="GI:732930"  
/db\_xref="SGD:50004865"  
/translation="MFGKVFVYIRTRTGPELSTIYTPVSSSLSDKCAECPFKKM  
HELMNSKOEFIORPVKNYRHQYDPSSTNNVSKLSTIGMBHNDSPVFGIIFYNDIWK  
SPKNOELGNNNDMSSGSRSHSPFKQULLQK"  
complement (10777..12021)  
/note="YM9920.07c, unknown, len: 414, CAI: 0.13, hydrophobic domains"

CDS

CDS

complement (start=1)  
/product="unknown"  
/protein\_id="CA88580.1"  
/db\_xref="GI:732931"  
/db\_xref="GOA:004835"  
/db\_xref="SGD:50004866"  
/db\_xref="UniProt/Swiss-Prot:Q04835"  
/translation="MNSVPKWKREKNTHLVSKEMNPTSLQLPSTTSIPKSSNS  
NEDPNVGNBTTLQRISSKYLKNIIGLVLTYSYFNSAMVSTKYLENDPDIDANDR  
QIKPLQILVAVIYIGLIIYWKISITSDVPRCKPVKRWLVKRGCTGFGVFGM  
YSLMYLTLISDAVLTFLAASLIRLSWILKERTFKVALSLLSLGGLVLIIVPSF  
LFGPELTDSSQIVSSDPKSLIATVGLMGVLSMCVYIIIRYIGKRAHAIIVS  
YFSLIATVFIQINTIPMKFOIPHSKQWILFQNLGVSGFIFOLLITMGIGRERAG



TITLE  
 JOURNAL  
 PUBMED  
 REFERENCES  
 AUTHORS  
 TITLE  
 JOURNAL  
 FEATURES  
 source  
 repeat region

Callon, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T.,  
 James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A.,  
 Kyes, S., Chan, M. S., Nene, V., Shilom, S. J., Suh, B., Peterson, J.,  
 Anguillo, S., Vaitea, M., Allen, J., Selegue, J., Haft, D.,  
 Mather, M. W., Pardey, A. B., Martin, D. M. A., Faircland, A. H.,  
 Fiumholz, M. M., Roos, D. S., Ralph, S. A., McFadden, G. I.,  
 Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C.,  
 Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M. and  
 Barrell, B.  
 Genome sequence of the human malaria parasite *Plasmodium falciparum*  
 Nature 419 (6906), 498-511 (2002)  
 12368864  
 2 (bases 1 to 257757)  
 Gardner, M. J.  
 Direct Submission  
 Submitted (13-SEP-2002) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA  
 Location/Qualifiers  
 1..257757  
 /organism="Plasmodium falciparum 3D7"  
 /mol\_type="genomic DNA"  
 /isolate="13D7"  
 /db\_xref="taxon:36329"  
 /chromosome="11"  
 257...381

repeat_region	257. .381
/rpt_type=tandem	
repeat_region	305. .373
/rpt_type=tandem	
repeat_region	763. .787
/rpt_type=tandem	
repeat_region	1304. .1333
/rpt_type=tandem	
gene	complement (<1434. .51994)
	/locus_tag="pp11_0070"
mRNA	/locus_tag="pp11_0070"
	/locus_tag="pp11_0070"
CDS	complement {join(1434. .1828, 1910. .1994)}
	/locus_tag="pp11_0070"
	/codon_start=1
	/product="hypothetical protein"
	/protein_id="AA035659.1"
	/db_xref="gi:23495997"
	/translation="MCMFVCIIECTEKKKKKKKKKLPIYIHKIDVNNKAEBQNFVPEIYVGHLEVEKQFPSTLYPVYTIINRDVTKHAAVVAERKRSLEFVLGIKAAIIYIVNKKQKMEBNIKISFSIITAVTNSMDETIPSTKVTLFLFISISTKIV"
repeat_region	1872. .1902
/rpt_type=tandem	
repeat_region	1926. .1960
/rpt_type=tandem	
repeat_region	2199. .2240
/rpt_type=tandem	
repeat_region	2384. .2542
/rpt_type=tandem	
repeat_region	2582. .2650
/rpt_type=tandem	
gene	complement (<2647. .>4074)
	/locus_tag="pp11_0071"
mRNA	complement (<2647. .4074)

```

mRNA      complement(<2647. .>4074)
           /locus_tag="PP11_0071"
           complement(2647..4074)
CDS        /locus_tag="PP11_0071"
           /codon_start=1
           /product="RuvB DNA helicase, putative"
           /protein_id="AA035660.1"
           /db_xref="GI:2349598"
           /translation="MEINISNTSSNNKKERYNIHSHKGLGVNTNYLHENEISLS
DEKSNMFPDSCGLVGQFARERASLPVLVLRQKLAGCCILAAPSSGSGSALAGI
SRELRNMRKPVFLSGSEVYSNFKIKTEVLELRPRSIRHKIEKELVVEGEVDWAVE
ENELCYSLNKAKQINAIITLKSVKSGKTLRLAPRIHQEIAERKIGDIVYIENNG
HVRVLRGDCNDYAKEDYIEFDEYSLPKSGVHKHKEAVQOISLHDIIDLANPVGEDL
AVSYLNRKPKTEKIRTEKRVLEINTVKNKELSSGLAEIIPGLYIDEAHMDIDERSY
LNALISPLAPLIMATNGICCTVGTGNTIPBHGIPVULDLRIIIRKFPYTLKEIVQ
ITLIPRQTEKINTREKRVNIAKTCOSGIPRNVIT

```





SOURCE	Danio rerio (zebrafish)
ORGANISM	Danio rerio
ANATOMY	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE	1 (bases 1 to 170295)
AUTHORS	Hunter,G.
JOURNAL	Direct Submission
COMMENT	Submitted (06-JUN-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail equitiles: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jan 6, 2004 this sequence version replaced gi:17805672.

zfish-help@sanger.ac.uk  
 On Jan 6, 2004 this sequence version replaced gi:37805672.  
 Genome Center  
 Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality  $\geq 30$ ); an attempt was made to resolve all sequencing problems, such as compressions and repeats: all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a VAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhongrong Bao and Sean Eddy, submitted), and those beginning 'dr' were identified by Rick Waterman (Stephen Johnson lab, Maennu). For further information see [http://www.sanger.ac.uk/Projects/D\\_rerio/fishmask.shtml](http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml) CH211-14AB6 is from a CHOR1-211 BAC library VECTOR: pTARBAC2.1.

FEATURES	Location/Qualifiers
source	1. .170295

```

/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-144B6"
/clone_1fb="CHOR1-211"
ORIGIN

```

Query Match	6.1%	Score 43.8;	DB 5;	length 170295;
Best Local Similarity	57.8%;	Pred. No. 0.3;		
Matches 78;	Conservative 0;	Mismatches 57;	Indels 0;	Gaps 0;

Qy 588 GGTGTTTTCATCATATAAAGGAGAAATTGGTCTCACTATGACAGTTTCGGG 647  
Db 60267 ATTTAATTATCAATTATAAATTATATTATTATTATTATTATTATTATTATTATGTG 60322

OY  
 648 ACGCTTAACCTTTATTCGACAGGAGCATCAATCAATCAGTATGTCAAAAAAAAAA 707  
 D6332 TTGTTATATTAATATTAAACACACACACACACTATTAATAATTAATATA 6038

Qy	708	AGACTAATAATACA	722
Db	60387	ATACTAATAATAATA	60401

RESULT	10				
AC092573/C					
LOCUS	AC092573	171265 bp	DNA	linear	PRI 01-MAR-2002
DEFINITION	Homo sapiens BAC clone Rpl1-107	from 2,	complete sequence.		
ACCESSION	AC092573	AC015764			
VERSION	AC092573.2	GI:15668084			
KEYWORDS	HTG.				
SOURCE	Homo sapiens (human)				

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PIRMEID  
9847074

1 (bases 1 to 171265)  
Sulston,J.E. and Waterston,R.  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)  
99063792

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE
2 (bases 1 to 171265)	Tomlinson, C., Cotton, M., Elliott, G., Dixon, R., Hawkins, M. and Boyer, E.	The sequence of Homo sapiens BAC clone RP11-107	unpublished (2001)	3 (bases 1 to 171265)

**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**  
  
 Waterston, R. H.  
 Direct Submission  
 Submitted (19-JUL-2001) Genome Sequencing Center Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 4 (bases 1 to 171265)  
 Waterston, R. H.

**TITLE** Direct Submission  
**JOURNAL** Submitted (19-SEP-2001) Genome Sequencing Center, Washington  
University School of Medicine, 444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
**REFERENCE** 5 (bases 1 to 171265)  
**AUTHORS** Waterston, R.  
**TITLE** Direct Submission

**JOURNAL** Submitted (01-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
**COMMENT** On Sep 19, 2001 this sequence version replaced gi:14916158.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC

```

-----
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics -----
Center project name: H.NH0001007
Drafting Center: WIBR
-----

```

**NOTICE:** This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., paired quality >= 30) ; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:** Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

**SOURCE INFORMATION:**  
The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frenken, E.,

Tateno, M., Catanese, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBAC3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-291G2, 1984 bp overlap; the clone sequenced to the right is RP11-158L8. Actual start of this clone is at base position 105122 of RP11-291G2; actual end is at base position 171265 of RP11-107.

Data from AC013461, AC073465, AC011667 and AC012052 was used to finish this clone, AC015764.

Polymorphisms have been identified between AC013461, AC015764 and AC012052.

The sequence of AC015764 has been incorporated into AC092573.

#### FEATURES

##### source

```
1.171265
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-107"
/clone_1kb="RPCT-11"
repeat_region
570..784
/rpc_family="MIR"
repeat_region
1544..1678
/rpc_family="Alu"
repeat_region
1678..1738
/rpc_family="(TRAG)n"
repeat_region
1813..1835
/rpc_family="AT_rich"
repeat_region
2054..2251
/rpc_family="AT_rich"
repeat_region
2173..2188
/rpc_family="Alu"
misc_feature
2173..2188
/notes="similar to Homo sapiens EST AL524174 (NID:g12787667)"
repeat_region
2352..2384
/rpc_family="L1"
repeat_region
2385..3107
/rpc_family="L1"
repeat_region
3117..3138
/rpc_family="AT_rich"
repeat_region
3120..3587
/rpc_family="MER1_type"
repeat_region
3588..3651
/rpc_family="MER1_type"
misc_feature
3877..3996
/notes="match to EST BE867724 (NID:g10316500)"
misc_feature
3947..3996
/notes="match to EST AL537636 (NID:g12801129)"
misc_feature
3956..3996
/notes="match to EST AW814670 (NID:g7907664)"
repeat_region
4224..4246
/rpc_family="AT_rich"
repeat_region
5266..5556
/rpc_family="Alu"
repeat_region
5997..6061
/rpc_family="CRI"
misc_feature
6244..6318
/notes="match to EST AL537636 (NID:g12801129)"
misc_feature
6244..6318
/notes="match to EST AW814670 (NID:g7907664)"
misc_feature
6244..6318
/notes="match to EST BE867724 (NID:g10316500)"
repeat_region
6547..6628
/rpc_family="L2"
repeat_region
6787..6814
```

```
repeat_region
6802..6873
/rpc_family="(TTCA)n"
repeat_region
6918..7214
/rpc_family="L2"
repeat_region
7234..7340
/rpc_family="Alu"
repeat_region
7499..7802
/rpc_family="MIR"
repeat_region
7961..8399
/rpc_family="Alu"
misc_feature
8787..9076
/notes="match to EST W25250 (NID:g1303104) zb68e09.r1"
misc_feature
8926..9076
/notes="similar to Bos taurus EST BF653485 (NID:g11918617)"
misc_feature
8926..9076
/notes="match to EST AL537636 (NID:g12801129)"
misc_feature
8926..9076
/notes="match to EST AW814670 (NID:g7907664)"
misc_feature
8926..9076
/notes="match to EST BE867724 (NID:g10316500)"
misc_feature
8999..9076
/notes="similar to Sus scrofa EST BE232128 (NID:g9016846)"
repeat_region
9173..9233
/rpc_family="MALR"
repeat_region
9234..9420
/rpc_family="MER1_type"
repeat_region
9421..9897
/rpc_family="MALR"
misc_feature
9698..9699
/rpc_family="MALR"
repeat_region
9935..10207
/notes="match to EST N95341 (NID:g1267676) zb68f09.s1"
repeat_region
10879..11046
/rpc_family="L1"
repeat_region
11288..11626
/rpc_family="MIR"
misc_feature
11288..11626
/notes="match to EST BF155794 (NID:g11050977)"
misc_feature
11231..11905
/notes="match to EST BF155794 (NID:g11050977)"
misc_feature
11231..11710
/notes="match to EST AL537636 (NID:g12801129)"
misc_feature
11231..11566
/notes="similar to Sus scrofa EST BE232128 (NID:g9016846)"
misc_feature
11231..11558
/notes="match to EST AW814670 (NID:g7907664)"
misc_feature
11231..11558
/notes="match to EST BE867724 (NID:g10316500)"
misc_feature
11231..11445
/notes="similar to Bos taurus EST BF653485 (NID:g11918617)"
misc_feature
11412..11967
/notes="similar to Bos taurus EST BF653485 (NID:g11918617)"
misc_feature
11617..11963
/notes="match to EST N95341 (NID:g1267676) zb68f09.s1"
misc_feature
11742..12650
/notes="match to EST AL347484 (NID:g4084690) tc06e04.x1"
misc_feature
11754..11994
/notes="match to EST AL537636 (NID:g12801129)"
misc_feature
11902..12561
/notes="similar to Homo sapiens EST BE161875 (NID:g8624596)"
misc_feature
11902..12561
/notes="similar to Homo sapiens EST BE793336 (NID:g1015317)"

Query Match
Best local Similarity 56.0%; Score 43.6; DB 9; Length 171265;
Matches 102; Conservative 0; Mismatches 79; Indels 1; Gaps 1;

QY 542 TTGATGAGAAATTAGCAAGCGGAAAAAACTATGCTAGCTGGAGTGTTCAT 601
Db 56108 TTAGTAAAAAAGAAAAAAGAAAAAAGCTGGCTTTGTACTCTGTTTGA 56049
QY 602 CATATAAGGAGAAATGTTGCTCATATGAGACAGTTCTGGAGAGCTTA-AC 660
Db 56048 GCTATAAAAAGGAAATTAATTCACACATCTCACACATTCGGAGAGAAA 55989
QY 661 TATTCAGAGAGCATATCATATCATATATTTGCAAAAAAAGAACTATATTA 720
Db 55988 TCTGAGACTAGGCTCTCAAGGCTCTTAAGAGCGAGAGGAAACAAAGAT 55929
```

```

QY          721 CA 722
Db          55928 CA 55927

RESULT 11
AC012052/c 208729 bp DNA linear HTG 04-MAY-2001
LOCUS      Homo sapiens chromosome 2 clone RP11-337F19 map 2, WORKING DRAFT
DEFINITION
AC012052
SEQUENCE   23 unordered pieces.
AC012052
AC012052.3 GI:8576227
HTG; HTGS PHASE1; HTGS _DRAFT.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE     Homo sapiens chromosome 2, clone RP11-337F19
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 208729)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
TITLE     Baldwin,J., Barna,A., Beckerly,R., Boguslavsky,L., Bouckgalter,B.,
JOURNAL   Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
REFERENCE Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
AUTHORS   Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
TITLE     Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
JOURNAL   Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
REFERENCE Lehocky,J., Lien,C., Locke,K., Macdonald,P., Marquis,N.,
AUTHORS   McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
TITLE     Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
JOURNAL   Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
REFERENCE Stange-Tromann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
AUTHORS   Teifaye,S., Tirelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
TITLE     Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
JOURNAL   Direct Submission
COMMENT    Submitted (19-OCT-1999) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
          On Jun 21, 2000 this sequence version replaced gi:7321517.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html

          ----- Genome Center
          Center: Whitehead Institute/ MIT Center for Genome Research
          Center code: WIBR
          Web site: http://www-seq.wi.mit.edu
          Contact: sequence.submissions@genome.wi.mit.edu

          ----- Project Information
          Center project name: L1385
          Center clone name: 337_F_19

          ----- Summary Statistics
          Sequencing vector: M13; M77815; 100% of reads
          Chemistry: Dye-terminator-amerham; 3% of reads
          Chemistry: Dye-terminator Big Dye; 97% of reads
          Assembly program: Phrap; version 0.960731
          Consensus quality: 191830 bases at least Q40
          Consensus quality: 199935 bases at least Q30
          Consensus quality: 203385 bases at least Q20
          Insert size: 210000; agarose-fp
          Insert size: 206529; sum-of-contigs
          Quality coverage: 4.0 in Q20 bases; agarose-fp
          Quality coverage: 4.1 in Q20 bas.

          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 23 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          1 1080: contig of 1080 bp in length

```

```

FEATURES
source
1..208729
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-337F19"
/clone_1b="RPCT-11 Human Male BAC"
1..1080
/note="assembly_fragment"
1181..2268
/note="assembly_fragment"
1181..2268
/note="assembly_fragment"
clone_end:T7
vector_side:right"
2369..3746
/note="assembly_fragment"
3847..6128
/note="assembly_fragment"
6229..7907
/note="assembly_fragment"
8008..10148
/note="assembly_fragment"
10249..12611
/note="assembly_fragment"
12712..14921
/note="assembly_fragment"
15022..18148
/note="assembly_fragment"
1081 1180: gap of 100 bp
1181 2268: contig of 1088 bp in length
2269 2368: gap of 100 bp
2369 3746: contig of 1378 bp in length
3747 3846: gap of 100 bp
3847 6128: contig of 2282 bp in length
6129 6228: gap of 100 bp
6229 7907: contig of 1679 bp in length
7908 8007: gap of 100 bp
8008 10148: contig of 2141 bp in length
10149 10248: gap of 100 bp
10249 12611: contig of 2363 bp in length
12612 12711: gap of 100 bp
12712 14921: contig of 2210 bp in length
14922 15021: gap of 100 bp
15022 18148: contig of 3127 bp in length
18149 18248: gap of 100 bp
18249 22510: contig of 4262 bp in length
22511 22511: gap of 100 bp
22512 27371: contig of 4761 bp in length
27372 27471: gap of 100 bp
27472 32405: contig of 4934 bp in length
32406 32505: gap of 100 bp
32506 38167: contig of 5662 bp in length
38168 38267: gap of 100 bp
38268 46372: contig of 8105 bp in length
46373 46472: gap of 100 bp
46473 56511: contig of 10079 bp in length
56512 56552: gap of 100 bp
56553 64715: contig of 8064 bp in length
64716 64815: gap of 100 bp
64816 75459: contig of 10644 bp in length
75460 75559: gap of 100 bp
75560 90744: contig of 15185 bp in length
90745 90844: gap of 100 bp
90845 107238: contig of 16394 bp in length
107239 128875: gap of 100 bp
128876 128975: gap of 100 bp
128976 152432: contig of 23447 bp in length
152433 152523: gap of 100 bp
152524 178410: contig of 25887 bp in length
178410 208729: gap of 100 bp
208729: contig of 30220 bp in length.
location/Qualifiers
1..208729
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-337F19"
/clone_1b="RPCT-11 Human Male BAC"
1..1080
/note="assembly_fragment"
1181..2268
/note="assembly_fragment"
1181..2268
/note="assembly_fragment"
clone_end:T7
vector_side:right"
2369..3746
/note="assembly_fragment"
3847..6128
/note="assembly_fragment"
6229..7907
/note="assembly_fragment"
8008..10148
/note="assembly_fragment"
10249..12611
/note="assembly_fragment"
12712..14921
/note="assembly_fragment"
15022..18148
/note="assembly_fragment"

```

```
repeat_region 4687.4829  
/rpt_family="MER5B"
```

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 195932)	Birren, B., Linton, L., Nusbaum, C. and Lander, E	Homo sapiens chromosome X. clone RP11-12D5



together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information on the WormPep database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/ChrX>

RP6-190D15 is from the library RPCI-6 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pPAC4

IMPORTANT: This sequence is not the entire insert of clone RP6-190D15. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP4-537K23 is at 28759 in this sequence. The true right end of clone RP4-753P9 is at 100 in this sequence.

## FEATURES

## source

```
1. .28858
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="q25-26.1"
/clone="RP6-190D15"
/clone_1b="RPCI-6"
39..338
/note="AluY repeat: matches 1..300 of consensus"
347..646
/note="AluSx repeat: matches 1..299 of consensus"
3076..3129
/note="27 copies 2 mer tc 77% conserved"
5914..6224
/note="AluYb8 repeat: matches 1..311 of consensus"
6287..6596
/note="AluSq repeat: matches 1..312 of consensus"
6597..8469
/note="L1MB8 repeat: matches 4267..6133 of consensus"
8470..8757
/note="AluSx repeat: matches 5..302 of consensus"
8758..8796
/note="L1MB8 repeat: matches 6133..6171 of consensus"
8818..9104
/note="AluSg1 repeat: matches 1..289 of consensus"
9345..9465
/note="L2 repeat: matches 2602..2728 of consensus"
10287..10362
/note="38 copies 2 mer aa 69% conserved"
10380..10462
/note="L1MB3 repeat: matches 6035..6127 of consensus"
10464..10518
/note="78K repeat: matches 1..53 of consensus"
10835..11133
/note="AluSx repeat: matches 1..298 of consensus"
11889..12080
/note="AluSg repeat: matches 24..296 of consensus"
12081..12122
/note="AluSg repeat: matches 253..294 of consensus"
12927..12974
/note="24 copies 2 mer gt 72% conserved"
13691..13842
repeat_region
```

```
repeat_region
/note="MIR repeat: matches 65..216 of consensus"
13957..14317
/note="L1M4 repeat: matches 4795..5162 of consensus"
14354..14450
/note="L1M2 repeat: matches 6213..6308 of consensus"
14610..14726
/note="L1M9 repeat: matches 6110..6218 of consensus"
14732..14837
/note="MER8 repeat: matches 49..156 of consensus"
14838..14908
/note="TIGGER2 repeat: matches 1..70 of consensus"
14909..15125
/note="MER8 repeat: matches 1..345 of consensus"
15171..15249
/note="TIGGER2 repeat: matches 1..73 of consensus"
15250..15315
/note="33 copies 2 mer aa 68% conserved"
15620..15923
/note="AluSq repeat: matches 1..312 of consensus"
15943..15956
/note="L1MB4 repeat: matches 5639..5652 of consensus"
15957..16210
/note="AluSg repeat: matches 38..290 of consensus"
16081..16168
/note="Weakly double-stranded"
16213..16724
/note="L1MB4 repeat: matches 5646..6176 of consensus"
16725..17110
/note="MIR1B repeat: matches 1..388 of consensus"
17111..17277
/note="L1MB4 repeat: matches 5487..5639 of consensus"
17640..17949
/note="AluY repeat: matches 1..309 of consensus"
18445..18745
/note="AluSx repeat: matches 1..302 of consensus"
19047..19348
/note="AluYb repeat: matches 1..291 of consensus"
19523..19744
/note="L1PA12 repeat: matches 5937..6158 of consensus"
19745..20039
/note="L1P4 repeat: matches 5345..5653 of consensus"
20040..20339
/note="AluY repeat: matches 1..301 of consensus"
20391..20689
/note="AluSg repeat: matches 1..299 of consensus"
20699..21369
/note="L1PAB repeat: matches 5489..6163 of consensus"
21430..21459
/note="U6 repeat: matches 1..30 of consensus"
21892..22025
/note="MIR repeat: matches 115..254 of consensus"
23037..23338
/note="AluSx repeat: matches 1..296 of consensus"
23378..23562
/note="L1M4 repeat: matches 5627..5818 of consensus"
23770..23985
/note="AluSx repeat: matches 1..216 of consensus"
23986..24027
/note="Alu repeat: matches 259..298 of consensus"
24570..24787
/note="L1MB repeat: matches 5500..5714 of consensus"
25661..25968
/note="AluSx repeat: matches 1..312 of consensus"
25970..26067
/note="49 copies 2 mer ag 61% conserved"
26552..26865
/note="AluSx repeat: matches 1..312 of consensus"
27019..27089
/note="MIR51B repeat: matches 453..526 of consensus"
27090..27396
/note="AluSc repeat: matches 1..306 of consensus"
27397..27415
/note="MIR51B repeat: matches 525..543 of consensus"
```



repeat\_region 27925..28083  
/note="12 repeat: matches 2535..2736 of consensus"  
repeat\_region 28662..28805  
/note="MIR repeat: matches 110..256 of consensus"  
ORIGIN

Query Match 5.8%; Score 42; DB 9; Length 28858;  
Best Local Similarity 59.0%; Pred.No. 0.77; Indels 0; Gaps 0;  
Matches 72; Conservative 0; Mismatches 50;  
QY 590 TTGTTTTCATCATTAAGGAGAAATTTGCTCATATGACAGTTTCTGGAC 649  
DB 13791 TTGTTTTCATCATTAAGGAGAAATTTGCTCATATGACAGTTTCTGGAC 13732  
QY 650 GTCTTACTTTATGCGAGAGACATCAATCATACAGATTTGCAAAAAAAG 709  
DB 13731 ATTGTCAGATCAATGAGATGTTATGAAAAATTTGAAAAAAGAAAAA 13672  
QY 710 AC 711  
DB 13671 AC 13670

RESULT 15  
AL645468/c 209764 bp DNA linear ROD 26-JUL-2002  
LOCUS AL645468  
DEFINITION Mouse DNA sequence from clone RP23-246F18 on chromosome 4, complete  
ACCESSION AL645468  
VERSION AL645468.11 GI:22002669  
KEYWORDS HTG.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 209764)  
Blakey, S.  
Direct Submission  
Submitted (26-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Jul 29, 2002 this sequence version replaced gi:21727348.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
-----

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; SW:  
SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep RP23-246F18 is  
from the RPI-23 Mouse PAC Library  
constructed by the group of Pieter de Jong.  
For further details see http://www.chori.org/bacpac/home.htm  
VECTOR: pBAC3.6.

FEATURES  
source  
1..209764  
Location/Qualifiers  
/organism="Mus musculus"

/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="4"  
/clone="RP23-246F18"  
/clone\_1ib="RPI-23"

ORIGIN  
Query Match 5.8%; Score 41.8; DB 10; Length 209764;  
Best Local Similarity 63.4%; Pred.No. 1.2; Indels 0; Gaps 0;  
Matches 64; Conservative 0; Mismatches 37;  
QY 127 TAAGAGTATTCATCTTCTTTTCCCTTCGCGGTCGAGACCCCTCTCCCC 186  
DB 70633 TAAAAATGTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCC 70574  
QY 187 GCACGATTTCTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCC 227  
DB 70573 TTCCCTTTCCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCC 70533

Search completed: February 7, 2005, 22:23:44  
Job time : 3606 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2005, 21:09:32 ; Search time 530 Seconds

(without alignments)  
8075.421 Million cell updates/sec

Title: US-10-776-213-2

Perfect score: 723

Sequence: 1 cttcgattacgcgcacac.....aaaagactataataacat 723

Scoring table: IDENTITY NUC  
Gapop 10-0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_15Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

. Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	723	100.0	723	4	AAD07475
2	723	100.0	11427	4	AAD07497
3	723	100.0	13073	4	AAD07493
4	720.4	99.6	850	4	AAD07503
5	497.4	68.8	680	6	ABQ76446
6	39	5.4	2000	8	ADA71938
7	38.6	5.3	2311	5	ABSA1325
8	38.4	5.3	544	11	ADT96256
9	37.8	5.2	7736	5	AAS29224
10	37.8	5.2	7736	5	ABAI6123
11	37.8	5.2	7736	6	ABSA68364
12	37.8	5.2	7736	10	ADC25486
13	37.4	5.2	386	5	ABV04394
14	37	5.1	366	4	AAL35477
15	37	5.1	366	8	ABX58465
16	37	5.1	366	12	ADJ28192
17	37	5.1	390	8	ABX36811
18	37	5.1	400	6	ABQ59188
19	36.4	5.0	201	13	ADS41207
20	36.4	5.0	591	13	ACN58226

21	36.4	5.0	110300	13	ADS36499	AdS36499 Human aut
22	36.2	5.0	524	10	ABT22885	ABT22885 Breast ca
23	36	5.0	7624	6	ABJ34113	ABJ34113 Human imm
24	35.8	5.0	553	13	ACN62578	ACN62578 Cotton de
25	35.8	5.0	815	2	AAZ00806	AAZ00806 Human sec
26	35.8	5.0	815	8	ADA39926	ADA39926 Human sec
27	35.8	5.0	815	8	ACCS0510	ACCS0510 Human sec
28	35.8	5.0	815	10	ADAS6114	ADAS6114 Gene enco
29	35.8	5.0	1143	8	ACA40062	ACA40062 Prokaryot
30	35.6	4.9	10517	13	ADS89438	ADS89438 Oligonuc
31	35.6	4.9	506	9	ACH17729	ACH17729 Human adu
32	35.6	4.9	5629	6	ABL58962	ABL58962 Human tum
33	35.6	4.9	5988	6	ABK09744	ABK09744 Human ova
34	35.6	4.9	5988	10	ADH29006	ADH29006 Human chr
35	35.6	4.9	5988	12	AD182483	AD182483 Human mod
36	35.6	4.9	64796	13	ACN37231	ACN37231 Human per
37	35.4	4.9	844	5	ABV18166	ABV18166 Human pro
38	35.4	4.9	33353	4	AAK70003	AAK70003 Human imm
39	35.4	4.9	201239	8	ACA64924	ACA64924 Human PLZ
40	35.2	4.9	281	6	ABV96856	ABV96856 Human pan
41	35.2	4.9	1837	2	AAZ41383	AAZ41383 Human nor
42	35.2	4.9	2000	8	ADA71938	ADA71938 Rice gene
43	35.2	4.9	2938	11	ACN88712	ACN88712 Breast ca
44	35.2	4.9	3197	13	ADS89390	ADS89390 Oligonuc
45	35.2	4.9	107330	12	ADQ97316	ADQ97316 Mouse can

## ALIGNMENTS

RESULT 1	
AAD07475	AAD07475 standard; DNA; 723 BP.
XX	XX
AC	AAD07475;
XX	XX
DT	10-AUG-2001 (first entry)
XX	XX
DE	Yeast promoter YMR251WA.
XX	XX
KW	Yeast; promoter; gene expression; fermentable carbon source; glucose;
KM	non-fermentable carbon source; ethanol; yeast cell culture; ds.
XX	XX
OS	Saccharomyces cerevisiae.
XX	XX
PN	W0200138549-A1.
XX	XX
PD	31-MAY-2001.
XX	XX
PF	17-NOV-2000; 2000MO-SE002277.
XX	XX
PR	23-NOV-1999; 99SE-00004247.
XX	XX
PA	(ASTR ) ASTRAZENECA AB.
XX	XX
PI	Belfield G, Oakley C;
XX	XX
DR	WPI; 2001-367697/38.
XX	XX
PT	New promoter sequences from Saccharomyces cerevisiae useful for
PT	controlling expression of homologous and heterologous nucleic acid
PS	expression in yeast cells.
XX	XX
PS	Claim 1; Page 67-68; 191pp; English.
XX	XX
CC	The invention relates to yeast promoters that are used to control the
CC	expression of homologous and heterologous nucleic acids encoding proteins
CC	and polypeptides in yeast cells. The yeast promoters are induced by a
CC	fermentable carbon source such as glucose or a non-fermentable carbon
CC	source such as ethanol or both. Therefore expression of nucleic acid
CC	molecules encoding a polypeptide under the control of the novel yeast
CC	promoters are regulated by varying the level of a fermentable carbon
CC	source or a non-fermentable carbon source or both. The yeast promoters

CC are useful for, inter alia, the high level production of proteins or  
 CC polypeptides in yeast cell culture. The present DNA sequence is  
 CC Saccharomyces cerevisiae YMR251AP promoter related to the invention  
 XX  
 SQ Sequence 723 BP; 199 A; 165 C; 166 G; 193 T; 0 U; 0 Other;

Query Match 100.0%; Score 723; DB 4; Length 723;  
 Best Local Similarity 100.0%; Pred. No. 9.2e-211;  
 Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CTTTCGATTAGACGACACATCATAGATGCGTCATTAATAATACACTACGAGAAA 60
DB 1 CTTTCGATTAGACGACACATCATAGATGCGTCATTAATAATACACTACGAGAAA 60
QY 61 ACCATAAGAGCAAGGATACCTACTTGGAGAGAAAGAGACGCTTGTAGGGGAT 120
DB 61 ACCATAAGAGCAAGGATACCTACTTGGAGAGAAAGAGACGCTTGTAGGGGAT 120
QY 121 GGGGGCTAAGAGATCTTCTTTCCCTTGGGCGGAGCCGGGACCCCTCCT 180
DB 121 GGGGGCTAAGAGATCTTCTTTCCCTTGGGCGGAGCCGGGACCCCTCCT 180
QY 181 CTCGCCGACGATTTCTTCTTTCATATCTTCTTTATTCCTATCCCGTTGAAGCAACC 240
DB 181 CTCGCCGACGATTTCTTCTTTCATATCTTCTTTATTCCTATCCCGTTGAAGCAACC 240
QY 241 GCACATAGACTAAATGGTGTGCGACATCTCATAGGCTGTGATCTGTATCTCACAGT 300
DB 241 GCACATAGACTAAATGGTGTGCGACATCTCATAGGCTGTGATCTGTATCTCACAGT 300
QY 301 GGTAAACGGACCGGTGCTCGGAAAACGGTCTTCTGTGACAAATCTAGAACAGGGGCTACA 360
DB 301 GGTAAACGGACCGGTGCTCGGAAAACGGTCTTCTGTGACAAATCTAGAACAGGGGCTACA 360
QY 361 GTCTCGATTATTAATTAATAAGGCAATTTTGTCTAGCGCGCGCGCGCGCTTCCCA 420
DB 361 GTCTCGATTATTAATTAATAAGGCAATTTTGTCTAGCGCGCGCGCGCGCTTCCCA 420
QY 421 ATAGGAGGCGGCAATTAATGAGGAGGCTCTACTCTTCTTATTTGGGTAAAGCCCTTTC 480
DB 421 ATAGGAGGCGGCAATTAATGAGGAGGCTCTACTCTTCTTATTTGGGTAAAGCCCTTTC 480
QY 481 TGTTCGCGGCAATGTTGCTGACGCTGCGCGGAGACATAGTATAGGATGTAC 540
DB 481 TGTTCGCGGCAATGTTGCTGACGCTGCGCGGAGACATAGTATAGGATGTAC 540
QY 541 TTTTCGATGAGAAATTAAGCAAGGCAAAAAAATCTATGCTAGGAGTTGTTTCA 600
DB 541 TTTTCGATGAGAAATTAAGCAAGGCAAAAAAATCTATGCTAGGAGTTGTTTCA 600
QY 601 TCATATTAAGAGGAGAAATTTGCTCACTATGTGACAGTTTCTGGAGCTTAACTTT 660
DB 601 TCATATTAAGAGGAGAAATTTGCTCACTATGTGACAGTTTCTGGAGCTTAACTTT 660
QY 661 TATTCGAGAGGACTATCAATTCATACAGATATTTGTCAAAAAAAGAGCTAATATTA 720
DB 661 TATTCGAGAGGACTATCAATTCATACAGATATTTGTCAAAAAAAGAGCTAATATTA 720
QY 721 CAT 723
DB 721 CAT 723

```

RESULT 2  
 AAD07497

ID AAD07497 standard; DNA; 11427 BP.

XX AAD07497;

XX 10-AUG-2001 (first entry)

DE PYMR251AP DNA plasmid.

XX

KW Yeast; promoter; gene expression; fermentable carbon source; glucose;  
 KW non-fermentable carbon source; ethanol; yeast cell culture;  
 KW PYMR251AP plasmid; ds.

OS Saccharomyces cerevisiae.  
 OS unidentified.  
 OS Chimeric.

PN WC200138549-A1.

PD 31-MAY-2001.

PF 17-NOV-2000; 2000MO-SR002277.

PR 23-NOV-1999; 99SE-00004247.

PA (ASTR ) ASTRAZENECA AB.

PI Belfield G, Oakley C;

DR WPI; 2001-367697/38.

PT New promoter sequences from Saccharomyces cerevisiae useful for  
 PT controlling expression of homologous and heterologous nucleic acid  
 PT expression in yeast cells.

PS Example 4; Page 133-143; 191p; English.

CC The invention relates to yeast promoters that are used to control the  
 CC expression of homologous and heterologous nucleic acids encoding proteins  
 CC and polypeptides in yeast cells. The yeast promoters are induced by a  
 CC fermentable carbon source such as glucose or a non-fermentable carbon  
 CC source such as ethanol or both. Therefore expression of nucleic acid  
 CC molecules encoding a polypeptide under the control of the novel yeast  
 CC promoters are regulated by varying the level of a fermentable carbon  
 CC source or a non-fermentable carbon source or both. The yeast promoters  
 CC are useful for, inter alia, the high level production of proteins or  
 CC polypeptides in yeast cell culture. The present sequence is PYMR251AP  
 CC plasmid related to the invention. This plasmid contains pBR1 and yeast  
 CC YMR251AP promoter

SQ Sequence 11427 BP; 3131 A; 2589 C; 2527 G; 3180 T; 0 U; 0 Other;

Query Match 100.0%; Score 723; DB 4; Length 11427;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-210;  
 Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CTTTCGATTAGACGACACATCATAGATGCGTCATTAATAATACACTACGAGAAA 60
DB 15 CTTTCGATTAGACGACACATCATAGATGCGTCATTAATAATACACTACGAGAAA 74
QY 61 ACCATAAGAGCAAGGATACCTACTTGGAGAGAAAGAGACGCTTGTAGGGGAT 120
DB 75 ACCATAAGAGCAAGGATACCTACTTGGAGAGAAAGAGACGCTTGTAGGGGAT 134
QY 121 GGGGGCTAAGAGATCTTCTTTCCCTTGGGCGGAGCCGGGACCCCTCCT 180
DB 135 GGGGGCTAAGAGATCTTCTTTCCCTTGGGCGGAGCCGGGACCCCTCCT 194
QY 181 CTCGCCGACGATTTCTTCTTTCATATCTTCTTTATTCCTATCCCGTTGAAGCAACC 240
DB 195 CTCGCCGACGATTTCTTCTTTCATATCTTCTTTATTCCTATCCCGTTGAAGCAACC 254
QY 241 GCACATAGACTAAATGGTGTGCGACATCTCATAGGCTGTGATCTGTATCTCACAGT 300
DB 255 GCACATAGACTAAATGGTGTGCGACATCTCATAGGCTGTGATCTGTATCTCACAGT 314
QY 301 GGTAAACGGACCGGTGCTCGGAAAACGGTCTTGTGACAAATTCAGAACAGGGGCTACA 360
DB 315 GGTAAACGGACCGGTGCTCGGAAAACGGTCTTGTGACAAATTCAGAACAGGGGCTACA 374
QY 361 GTCTCGATTATTAATTAATAAGGCAATTTTGTCTAGCGCGCGCGCGCGCTTCCCA 420

```

Db 375 GTCTCATATATAGAAATTAATTAAGCCATTCTTTGCTAGCGCCGCCGCCGCTTTTCCA 434  
 Qy 421 ATAGGAGGCGCGAGTTTATCGCGGAGCTCTACTTCTTCTATTGGTAAGCCCTTTTC 480  
 Db 435 ATAGGAGGCGCGAGTTTATCGCGGAGCTCTACTTCTTCTATTGGTAAGCCCTTTTC 494  
 Qy 481 TGTTCCTGCGCGAGTGTCTGTCGAGGCTGCGCGGAGCAATATGTATTAAGGATGTAC 540  
 Db 495 TGTTCCTGCGCGAGTGTCTGTCGAGGCTGCGCGGAGCAATATGTATTAAGGATGTAC 554  
 Qy 541 TTTTCATGAGAGATTAAGCAAGCGGAAAAAACTATAGGCTAGCTGGAGTGTCTTCA 600  
 Db 555 TTTTCATGAGAGATTAAGCAAGCGGAAAAAACTATAGGCTAGCTGGAGTGTCTTCA 614  
 Qy 601 TCATATTAAGGAGAAATTTGTTCTCACTATGTGACAGTTTCTGGACGCTTTAACTTT 660  
 Db 615 TCATATTAAGGAGAAATTTGTTGCTCACTATGTGACAGTTTCTGGACGCTTTAACTTT 674  
 Qy 661 TATTGCAAGAGACTATCAATCATACATATTTGTCAAAAAAAGACTAATATTA 720  
 Db 675 TATTGCAAGAGACTATCAATCATACATATTTGTCAAAAAAAGACTAATATTA 734  
 Qy 721 CAT 723  
 Db 735 CAT 737

## RESULT 3

AAD07493  
 ID AAD07493 standard; DNA; 13073 BP.

AC AAD07493;

DE 10-AUG-2001 (first entry)

XX pYMR251AP+luc sequence DNA construct.

XX Yeast; promoter; gene expression; fermentable carbon source; glucose;

KW non-fermentable carbon source; ethanol; yeast cell culture;

KM pYMR251AP plasmid; luciferase gene; ds.

XX Saccharomyces cerevisiae.

OS unidentified.

OS Chimeric.

PN WO200138549-A1.

PD 31-MAY-2001.

PF 17-NOV-2000; 2000MO-SE002277.

PR 23-NOV-1999; 99SE-00004247.

XX (ASTR) ASTRAZENECA AB.

PI Belfield G, Oakley C;

DR WPI; 2001-367697/38.

XX New promoter sequences from Saccharomyces cerevisiae useful for

PT controlling expression of homologous and heterologous nucleic acid

PT expression in yeast cells.

XX Example 3; Page 88-99; 191pp; English.

XX The invention relates to yeast promoters that are used to control the

XX expression of homologous and heterologous nucleic acids encoding proteins

XX and polypeptides in yeast cells. The yeast promoters are induced by a

XX fermentable carbon source such as glucose or a non-fermentable carbon

XX source such as ethanol or both. Therefore expression of nucleic acid

XX molecules encoding a polypeptide under the control of the novel yeast

XX promoters are regulated by varying the level of a fermentable carbon

XX source or a non-fermentable carbon source or both. The yeast promoters

CC are useful for, inter alia, the high level production of proteins or  
 CC polypeptides in yeast cell culture. The present sequence is pYMR251AP+luc  
 CC sequence DNA construct related to the invention. The pYMR251AP+luc  
 CC sequence contains pPRB1 plasmid, luc backbone (luciferase gene) and yeast  
 CC YMR251WA promoter  
 XX  
 SQ Sequence 13073 BP; 3584 A; 2949 C; 2934 G; 3606 T; 0 U; 0 Other;

Query Match 100.0%; Score 723; DB 4; Length 13073;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-210;  
 Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTCATTTAGCAGCGACACATCATAGAGTGGCTCATATAAATACACTCGGAAAA 60  
 Db 16 CTTTCATTTAGCAGCGACACATCATAGAGTGGCTCATATAAATACACTCGGAAAA 75  
 Qy 61 ACCATAAGAGCAAGACGATACCTACTTGGAGAGAGAGAGAGAGAGAGAGAGAGAT 120  
 Db 76 ACCATAAGAGCAAGACGATACCTACTTGGAGAGAGAGAGAGAGAGAGAGAGAT 135  
 Qy 121 GGGGGCTAAGAGTCAATTCATTTCTTTCCCTCGCGGTCCGAGCCCGGACCCCTCCT 180  
 Db 136 GGGGGCTAAGAGTCAATTCATTTCTTTCCCTCGCGGTCCGAGCCCGGACCCCTCCT 195  
 Qy 181 CTCCCGGACGATTTCTTCTCTCATATCTTCTTTATTTCCATTCCTGTAAGCAACC 240  
 Db 196 CTCCCGGACGATTTCTTCTCTCATATCTTCTTTATTTCCATTCCTGTAAGCAACC 255  
 Qy 241 GCACATAGACTAAGTGTGCTGAGCATCTCAATGCGTGTGATCTTGATGTACAGT 300  
 Db 256 GCACATAGACTAAGTGTGCTGAGCATCTCAATGCGTGTGATCTTGATGTACAGT 315  
 Qy 301 GGTAAAGGAGCAGCTGCTCGAAAAAGGTTCTTCTGTGACAAATCTAAGAGGCTTACA 360  
 Db 316 GGTAAAGGAGCAGCTGCTCGAAAAAGGTTCTTCTGTGACAAATCTAAGAGGCTTACA 375  
 Qy 361 GTCTCATTAATAGAAATTAATTAAGGAGCAATTTTGTATACGCGCGCGCGCGCTTCCCA 420  
 Db 376 GTCTCATTAATAGAAATTAATTAAGGAGCAATTTTGTATACGCGCGCGCGCGCTTCCCA 435  
 Qy 421 ATAGGAGGCGCGAGTTTATCGCGGAGCTCTACTTCTTCTATTGGTAAGCCCTTTC 480  
 Db 436 ATAGGAGGCGCGAGTTTATCGCGGAGCTCTACTTCTTCTATTGGTAAGCCCTTTC 495  
 Qy 481 TGTTCCTGCGCGAGTGTCTGTCGAGGCTGCGCGGAGCAATATGTATTAAGGATGTAC 540  
 Db 496 TGTTCCTGCGCGAGTGTCTGTCGAGGCTGCGCGGAGCAATATGTATTAAGGATGTAC 555  
 Qy 541 TTTTCATGAGAGATTAAGCAAGCGGAAAAAACTATAGGCTAGCTGGAGTGTCTTCA 600  
 Db 556 TTTTCATGAGAGATTAAGCAAGCGGAAAAAACTATAGGCTAGCTGGAGTGTCTTCA 615  
 Qy 601 TCATATTAAGGAGAAATTTGTTCTCACTATGTGACAGTTTCTGGACGCTTTAACTTT 660  
 Db 616 TCATATTAAGGAGAAATTTGTTCTCACTATGTGACAGTTTCTGGACGCTTTAACTTT 675  
 Qy 661 TATTGCAAGAGACTATCAATCATACAGATATTTGTCAAAAAAAGACTAATATTA 720  
 Db 676 TATTGCAAGAGACTATCAATCATACAGATATTTGTCAAAAAAAGACTAATATTA 735  
 Qy 721 CAT 723  
 Db 736 CAT 738

## RESULT 4

AAD07503  
 ID AAD07503 standard; DNA; 850 BP.

XX AAD07503;

DT 10-AUG-2001 (first entry)

DE	Yeast YMR251WA promoter region.
XX	
KM	Yeast; promoter; gene expression; fermentable carbon source; glucose;
XX	non-fermentable carbon source; ethanol; yeast cell culture; ds.
XX	
OS	Saccharomyces cerevisiae.
XX	
PH	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	1..64
FT	/tag= a
FT	/product= "Yeast YMR251W open reading frame encoding
FT	polypeptide"
FT	complement(78..95)
FT	/tag= c
FT	/bound moiety= "YMR251WA forward PCR primer"
FT	/partial
FT	786..812
FT	/tag= d
FT	/bound moiety= "YMR251WA reverse PCR primer"
FT	/partial
FT	801..850
FT	/tag= b
FT	/product= "Yeast YMR251WA open reading frame encoding
FT	polypeptide"
PN	WO200138549-A1.
XX	
PD	31-MAY-2001.
XX	
PF	17-NOV-2000; 2000WO-SE002277.
XX	
PR	23-NOV-1999; 99SE-00004247.
XX	
PA	(ASTR ) ASTRAZENECA AB.
PI	
XX	Belfield G, Oakley C;
DR	WPI, 2001-367697/38.
XX	
XX	
PT	New promoter sequences from <i>Saccharomyces cerevisiae</i> useful for
PT	controlling expression of homologous and heterologous nucleic acid
PT	expression in yeast cells.
PS	
XX	Example 3; Fig 14; 191pp; English.
XX	
CC	The invention relates to yeast promoters that are used to control the
CC	expression of homologous and heterologous nucleic acids encoding proteins
CC	and polypeptides in yeast cells. The yeast promoters are induced by a
CC	fermentable carbon source such as glucose or a non-fermentable carbon
CC	source such as ethanol or both. Therefore expression of nucleic acid
CC	molecules encoding a polypeptide under the control of the novel yeast
CC	promoters are regulated by varying the level of a fermentable carbon
CC	source or a non-fermentable carbon source or both. The yeast promoters
CC	are useful for, inter alia, the high level production of proteins or
CC	polypeptides in yeast cell culture. The present DNA sequence is
CC	<i>Saccharomyces cerevisiae</i> YMR251WA promoter region related to the
CC	invention
XX	
XX	
SQ	Sequence 850 BP; 225 A; 199 C; 193 G; 233 T; 0 U; 0 Other;
	Query Match 99.6%; Score 720.4; DB 4; Length 850;
	Best Local Similarity 99.9%; Pred. No. 6.2e-210;
	Matches 721; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	1 CTTTGATTTACACGACACACATACATACATGACCTGCGTCATTAATAATACACTACGAAAA 60
DB	78 CTTTGATTTACACGACACACATACATACATGACCTGCGTCATTAATAATACACTACGAAAA 137
QY	61 ACCATAAAGACCAAGGATGCTCTACTTGAAGAAAGAGACAGCTGTGTAAGGGGAT 120
DB	138 ACCATAAAGACCAAGGATGCTCTACTTGAAGAAAGAGACAGCTGTGTAAGGGGAT 197
Y	121 GGGGGCTTAAGAAGTCATTCCTTTTCCCTTCGCGGTCCGACCCGGGACCCCTCT 180

[illegible]









PR 14-SEP-2000; 2000US-023239P.  
PR 14-SEP-2000; 2000US-023240P.  
PR 14-SEP-2000; 2000US-023241P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246529P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246533P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249267P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 01-DEC-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.

PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251865P.  
PR 08-DEC-2000; 2000US-0251866P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Baraeh SC, Ruben SM,  
XX WPI; 2001-465557/50.  
XX  
XX Nucleic acid molecules encoding human secreted chromosomal binding  
XX proteins, used in preventing, treating or ameliorating a disorder, e.g.  
XX Alzheimer's and Parkinson's diseases and cancers.  
XX  
XX Disclosure; SEQ ID NO 333; 561pp; English.

XX The present invention relates to the isolation of novel DNA-binding  
XX proteins (AAU18154-AAU18281), and cDNA and genomic sequences encoding for  
XX these proteins. DNA-binding, proteins such as histones, chromo (chromatin  
XX organisation modifier) domain proteins, and Y-box binding proteins may  
XX contribute to diseases resulting from aberrant DNA organisation and/or  
XX gene transcription. The sequences of the invention are useful in  
XX screening assays to identify antagonists and/or agonists that may enhance  
XX or block activities mediated by DNA-binding proteins. Blockers of DNA-  
XX binding proteins may be useful in treating disorders such as malignant  
XX diseases (e.g. cancer), autoimmune disorders (e.g. diabetes mellitus),  
XX rheumatic diseases (e.g. rheumatoid arthritis), genetic abnormalities  
XX (e.g. cystic fibrosis), infectious diseases (e.g. HIV) and neurological  
XX disorders (e.g. Alzheimer's disease). The polynucleotide sequences of the  
XX invention may also be used in gene therapy. AAS29158-AAS29239 represent  
XX genomic sequences encoding for novel DNA-binding proteins. Note: The  
XX sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 7736 BP; 1672 A; 1981 C; 2164 G; 1919 T; 0 U; 0 Other;

Query Match 5.2%; Score 37.8; DB 5; Length 7736;  
Best Local Similarity 54.7%; Pred. No. 2.9;

Matches 75; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 110 GTAAGGGGATGGGGGCTAAGAGTCACTTCTTCTTCCCTTGGCGTCCGACCG 169  
DB 5184 GTTCAGGGGATGGAGTAAAGACACACAGTGTTCCTCCCCACAGCCGCGAGATGTG 5125  
QY 170 GGACCCCTCTCTCCCGCAGAGTTCTCTTCATATCTCTCTTATTCATCCG 229  
DB 5124 GAAGTACTCTCACTCTCCCGAGCTGCTTTCCTCATGCGCTGACCTCGCTCCC 5065  
QY 230 TTGAAGCAACCGCACTA 246  
DB 5064 TGGTAGCAGCTGTACCA 5048

RESULT 10  
ABA16123/c  
ABA16123 standard; DNA; 7736 BP.

XX ABA16123;

XX 23-JAN-2002 (first entry)

DE Human nervous system related polynucleotide SEQ ID NO 8454.

KW Human; nootropic; neuroprotective; cytosstatic; dermatological; vituicide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antischizking; antianaemic; antiarthritic; cancer;

KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW anti allergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotoxic; gene therapy; vaccine; ds.  
OS Homo sapiens.  
XX  
XX WO200159063-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001334.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225477P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231422P.  
PR 08-SEP-2000; 2000US-0231423P.  
PR 08-SEP-2000; 2000US-0231444P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.

PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 20-OCT-2000; 2000US-0242212P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 01-DEC-2000; 2000US-0251160P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256179P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251858P.  
PR 08-DEC-2000; 2000US-0251859P.  
PR 08-DEC-2000; 2000US-0251899P.

PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-541565/60.  
XX  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system cancers  
PT and metastases.  
XX  
XX  
PS Disclosure; SEQ ID NO 8454; 1701bp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (ABAI1004-ABA21534) and proteins  
CC (ABAI4678-ABAI8001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 7736 BP; 1672 A; 1981 C; 2164 G; 1919 T; 0 U; 0 Other;  
XX  
Query Match 5.2%; Score 37.8; DB 5; Length 7736;  
Best Local Similarity 54.7%; Pred. No. 2.9;  
Matches 75; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
XX  
QY 110 GTAAGGGGATGGGGGCTAAGAGTCATTCATTTCTTTCCCTTCGCGTCCGACCCG 169  
Db 5184 GTTCAGGGGATGGATGTAAAGACACACACAGTTGCCGCCGACGCCGAGATGTG 5125  
QY 170 GGACCCCTCTCTCCCGCAGATTTCTTCTTCAATCTCTTTTATTCATATCCG 229  
Db 5124 GAGATCTCACTCTCTCCGAGTCGCTTTCCATGAGCCTTGACCTCCCTCCC 5065  
QY 230 TTGAAGCAGCGCACTA 246  
Db 5064 TGGTAGCAGCTGTACCA 5048  
XX  
RESULT 11  
ABSE68364/c  
ID ABS68364 standard; DNA; 7736 BP.  
XX  
XX ABS68364;  
XX  
XX 18-NOV-2002 (first entry)  
XX  
XX Human DNA-binding protein genomic DNA sequence #67.  
XX  
XX Human; DNA-binding protein; B cell immunodeficiency; autoimmune disorder;  
XX severe combined immunodeficiency; rheumatoid arthritis; Crohn's disease;  
XX diabetes mellitus; allergy; asthma; inflammatory condition; thrombosis;  
XX graft-versus-host disease; blood-related disorder; atherosclerosis;  
XX hyperproliferative disorder; cancer; renal disorder; arhythmia;  
XX acute glomerulonephritis; cardiovascular disorder; respiratory disorder;  
XX Goodpasture's syndrome; neurological disorder; Alzheimer's disease;  
XX Parkinson's disease; endocrine disorder; Addison's disease;  
XX reproductive system disorder; endometriosis; infectious disease;  
XX viral infection; bacterial infection; fungal infection; vaccine;  
XX

KW gastrointestinal disorder; multiple sclerosis; gene therapy; ds.  
XX Homo sapiens.  
XX  
XX US2002102638-A1.  
XX  
XX 01-AUG-2002.  
XX  
XX 17-JAN-2001; 2001US-00764846.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
XX 14-AUG-2000; 2000US-0225758P.  
XX 22-AUG-2000; 2000US-0226688P.  
XX 30-AUG-2000; 2000US-0228924P.  
XX 01-SEP-2000; 2000US-0229287P.  
XX 01-SEP-2000; 2000US-0229343P.  
XX 01-SEP-2000; 2000US-0229344P.  
XX 01-SEP-2000; 2000US-0229345P.  
XX 05-SEP-2000; 2000US-0229509P.  
XX 05-SEP-2000; 2000US-0229513P.  
XX 08-SEP-2000; 2000US-0231413P.  
XX 21-SEP-2000; 2000US-0234223P.  
XX 21-SEP-2000; 2000US-0234274P.  
XX 25-SEP-2000; 2000US-0234597P.  
XX 27-SEP-2000; 2000US-0235834P.  
XX 29-SEP-2000; 2000US-0236327P.  
XX 29-SEP-2000; 2000US-0236367P.  
XX 29-SEP-2000; 2000US-0236368P.  
XX 29-SEP-2000; 2000US-0236369P.  
XX 29-SEP-2000; 2000US-0236370P.  
XX 29-SEP-2000; 2000US-0236802P.  
XX 02-OCT-2000; 2000US-0237037P.  
XX 02-OCT-2000; 2000US-0237038P.  
XX 02-OCT-2000; 2000US-0237039P.  
XX 02-OCT-2000; 2000US-0237040P.  
XX 13-OCT-2000; 2000US-0239335P.  
XX 20-OCT-2000; 2000US-0240960P.  
XX 20-OCT-2000; 2000US-0241785P.  
XX 20-OCT-2000; 2000US-0241809P.  
XX 01-NOV-2000; 2000US-0244617P.  
XX 17-NOV-2000; 2000US-0249299P.  
XX 08-DEC-2000; 2000US-0251556P.  
XX 08-DEC-2000; 2000US-0251688P.  
XX 08-DEC-2000; 2000US-0251689P.  
XX  
XX (ROSE/) ROSEN C A.  
XX (RUBE/) RUBEN S M.  
XX (BARA/) BARASH S C.  
XX  
XX Rosen CA, Ruben SM, Barash SC;  
XX WPI; 2002-690611/74.  
XX  
XX Novel DNA-binding protein useful for diagnosis, prognosis, prevention and  
XX treatment of immune, hyperproliferative, respiratory, cardiovascular,  
XX reproductive, endocrine, gastrointestinal and neurological disorders.  
XX

P5 Claim 1, SEQ ID NO 333; 2252pp; English.

XX The present invention relates to a new DNA-binding protein. The invention  
CC is useful in treating, preventing, diagnosing and/or prognosing  
CC immunodeficiencies (e.g. B cell immunodeficiencies), severe combined  
CC immunodeficiencies), autoimmune disorders (rheumatoid arthritis, multiple  
CC sclerosis, diabetes mellitus), allergic reactions and conditions (e.g.  
CC asthma), inflammatory conditions, graft-versus-host disease, blood-  
CC related disorders (thrombosis, atherosclerosis), hyperproliferative  
CC disorders (e.g. cancer), renal disorders (e.g. acute glomerulonephritis),  
CC cardiovascular disorders (e.g. arrhythmia), respiratory disorders  
CC (Goodpasture's syndrome), neurological disorders (e.g. Alzheimer's  
CC disease, Parkinson's disease), endocrine disorders (e.g. Addison's  
CC disease), reproductive system disorders (e.g. endometriosis), infectious  
CC diseases (e.g. viral, bacterial or fungal infections) and  
CC gastrointestinal disorders (e.g. Crohn's disease). The invention is also  
CC useful to stimulate neuronal growth and treat, prevent, and/or diagnose  
CC neuronal damage which occurs in certain neuronal disorders or neuro-  
CC degenerative conditions. The present nucleic acid sequence represents a  
CC human DNA-binding protein genomic DNA sequence of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from USPRO  
CC at <http://seqdata.uspto.gov/sequence>

SQ Sequence 7736 BP; 1672 A; 1981 C; 2164 G; 1919 T; 0 U; 0 Other;

Query Match 5.2%; Score 37.8; DB 6; Length 7736;  
Best Local Similarity 54.7%; Pred. No. 2.9;  
Matches 75; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

OY 110 GTAAGGGGATGGGGGCTTAAGAAGCATTTCACTTTCCTTCCTCGGGGTCCGACCCG 169  
DB 5184 GTTCAGGGGAATGAGATGTAAGACACACAGTTGTTCCCCCAGCGGCCCAATGTG 5125  
OY 170 GGACCCCCTCTCTCCCGCACGATTTCCTCTTATATATCTCTTATTCATTCGCCG 229  
DB 5124 GAAGTACTCTCACTCTCTCCCGAGTCTGCCTTTCCCTATGAGCTCTAAGCTCGTCTCC 5065  
OY 230 TTGAAGCAACCGCACTA 246  
DB 5064 TGGTAGCAGCTGTACC 5048

RESULT 12  
ADC25486/c  
ID ADC25486 standard; cDNA; 7736 BP.  
XX  
XX ADC25486;  
XX  
XX 18-DEC-2003 (first entry)  
DT  
XX  
DE Human cDNA from extracellular matrix gene 78 #3.  
XX  
KM Extracellular matrix protein; cytosolic; antibacterial; virucide;  
KM neuroprotective; gynaecological; gastrointestinal-gen; cardiant;  
KM cardiovascular-gen; nephrotoxic; antiinflammatory; muscular-gen;  
KM respiratory-gen; immunosuppressive; cerebroprotective; vasotropic;  
KM neotrophic; antiallergic; cancer; bacterial infection; viral infection;  
KM neural disorder; immune system disorder; blood disorder;  
KM muscular disorder; reproductive disorder; gastrointestinal disorder;  
KM pulmonary disorder; cardiovascular disorder; renal disorder;  
KM inflammatory disorder; proliferative disorder; human; gene therapy; ss;  
gene.  
XX  
OS Homo sapiens.  
XX  
FN US2003049650-A1.  
XX  
PD 13-MAR-2003.  
PP  
PF 07-MAR-2002; 2002US-00091483.  
PR 31-JAN-2000; 2000US-0179065P.

PR	02-OCT-2000;	2000US-0237038P.
PR	02-OCT-2000;	2000US-0237039P.
PR	02-OCT-2000;	2000US-0237040P.
PR	13-OCT-2000;	2000US-0239933P.
PR	13-OCT-2000;	2000US-0239937P.
PR	20-OCT-2000;	2000US-0240960P.
PR	20-OCT-2000;	2000US-0241221P.
PR	20-OCT-2000;	2000US-0241785P.
PR	20-OCT-2000;	2000US-0241786P.
PR	20-OCT-2000;	2000US-0241787P.
PR	20-OCT-2000;	2000US-0241808P.
PR	20-OCT-2000;	2000US-0241809P.
PR	01-NOV-2000;	2000US-0241826P.
PR	01-NOV-2000;	2000US-0244617P.
PR	08-NOV-2000;	2000US-0246474P.
PR	08-NOV-2000;	2000US-0246475P.
PR	08-NOV-2000;	2000US-0246476P.
PR	08-NOV-2000;	2000US-0246477P.
PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	08-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249246P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251688P.
PR	05-DEC-2000;	2000US-0256718P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251899P.
PR	08-DEC-2000;	2000US-0251907P.
PR	11-DEC-2000;	2000US-0254907P.
PR	05-JAN-2001;	2001US-0259678P.
PR	17-JAN-2001;	2001US-0076484F.
XX	(HUMA-) HUMAN GENOME SCI INC.	
XX	Rosen CA, Ruben SM, Barash SC,	
PI	WPI; 2003-605749/57.	
DR	P-PSDB; ADC25241.	
XX	New DNA-binding proteins and gene encoding them, useful for diagnosing	
PT	treating and/or preventing e.g. neurological, inflammatory, infectious	

PT cardiovascular, autoimmune, respiratory, neoplastic or digestive  
PT diseases.  
XX  
XX  
PS  
PS  
XX Disclosure; SEQ ID NO 333; 226pp; English.

CC The invention relates to an isolated nucleic acid molecule (cDNA)  
CC encoding a human extracellular matrix protein, representing one of 161  
CC novel genes. Also included are recombinant vectors, host cells  
CC (expressing the protein), the extracellular matrix proteins (including  
CC their fragments, epitopes and homologues), an isolated antibody that  
CC binds specifically to the protein, diagnosing a pathological condition or  
CC susceptibility to a pathological condition (comprising determining the  
CC presence or absence of a mutation in the nucleic acid and diagnosing a  
CC condition based on the presence or absence of the mutation), diagnosing a  
CC pathological condition or susceptibility to a pathological condition  
CC (comprising determining the presence or amount of expression of the  
CC protein in a biological sample and diagnosing a condition based on the  
CC presence or amount of expression of the protein), preventing, treating or  
CC ameliorating a medical condition by administering the nucleic acid or  
CC protein to a mammalian subject, identifying a binding partner to the  
CC protein, the gene corresponding to the cDNA sequence, and identifying an  
CC activity in a biological assay (comprising expressing the nucleic acid in  
CC a cell, isolating the supernatant, detecting an activity in a biological  
CC assay and identifying the protein in the supernatant having the  
CC activity). The nucleic acids and proteins display the following  
CC activities: Cytostatic, antibacterial, Virucide, Neuroprotective,

Query Match	5.2%	Score 37.8;	DB 10;	Length 7736;
Best Local Similarity	54.7%;	Pred. No. 2.9;		
Matches 75; Conservative	0;	Mismatches 62;	Indels 0;	Gaps 0;

QY	110	GTAAGGGGGATAGGGGGCTAAAGATCATATCACTTTCTTTTCCCTTGGCGGTCCGAGCCG	169
Db	5184	GTTTCAAGGGGATGATGTATTAAGCACAACACAGCTGTGTTCCCCCAAGCGGCCAGATGTG	5125
QY	170	GGACCCCTCCTCTCTCCCGCAGATTTCTTCCTTCAATATCTTCCTTTATTCATCCG	229
Db	5124	GAATATATTCACATCTCTCTCCCGAGCTGACCTTTTCCCTCATGGCCCTGAGACTCGCTCCC	5065
QY	230	TTGAAGCAACCGCACTA	246
Db	5064	TGCTAGCAGCTGTACCA	5048

RESULT	13	
ABV04394		
ID	ABV04394	standard; cDNA; 386 BP.
XX		
AC	ABV04394;	
XX		
DT	13-SEP-2002	(first entry)
XX		
DE	Human prostate expression marker	cDNA 4385.
XX		
KM	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;	pharmacogenomic marker; gene; ss.
XX		
OS	Homo sapiens.	
XX		
PN	WO200160860-A2.	
XX		
PD	23-AUG-2001.	
XX		
PF	20-FEB-2001; 2001WO-US005171.	
XX		
PR	17-FEB-2000; 2000US-0183119P.	
PR	16-MAR-2000; 2000US-0189862P.	
PR	25-MAY-2000; 2000US-0207544P.	
PR	09-JUN-2000; 2000US-0211314P.	
PR	18-JUL-2000; 2000US-0219007P.	
PR	13-DEC-2000; 2000US-0255281P.	
XX		
XX		

XX Schlegel R, Endege WO, Monahan JE;  
XX WPI; 2001-662795/76.  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer.  
XX Claim 1; Page 761; 11750pp; English.  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABY00010-ABY62213) of the  
CC specification or its complement. (I) is useful for: (a) assessing whether  
CC a patient is afflicted with prostate cancer; (b) monitoring the  
CC progression of prostate cancer in a patient; (c) assessing the efficacy  
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient  
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX  
SQ Sequence 386 BP; 108 A; 74 C; 74 G; 128 T; 0 U; 2 Other;  
Query Match 5.2%; Score 37.4; DB 5; Length 386;  
Best Local Similarity 64.4%; Pred. No. 0.95;  
Matches 56; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
QY 628 ACTATGTGACAGTTCGGACGCTTAATCTTATTCGAGAGACTATCAATCATACA 687  
DB 175 ATATATGGAGACCTCAGCTATTTTCAGGTCCTCTTAAGTACTAAGGAAACATCCCA 234  
QY 688 GATATGTGCAAAAAAAGACTTA 714  
DB 235 GATACCTTCANAAAAAAGTCTTA 261  
RESULT 14  
AAL35477/c  
ID AAL35477 standard; cDNA; 366 BP.  
XX AAL35477;  
XX  
DT 08-JAN-2002 (first entry)  
XX  
DE Human musculoskeletal system related polynucleotide SEQ ID NO 819.  
XX  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiatherogenic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein;  
KW musculoskeletal system; ss.  
XX  
XX Homo sapiens.  
XX  
PN WO200155367-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-US001338.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234497P.  
PR 25-SEP-2000; 2000US-0234988P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 27-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241212P.  
PR 20-OCT-2000; 2000US-0241785P.





11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218280P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 22-AUG-2000; 2000US-0225758P.  
PR 30-AUG-2000; 2000US-0226688P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234977P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
XX  
XX  
PA (ROSE/) ROSEN C A.  
PA (RUBE/) RUBEN S M.  
PA (BARA/) BARASH S C.  
PI Rosen CA, Ruben SM, Barash SC;  
XX  
XX  
DR WPI: 2003-128199/12.  
XX P-PSDB; ABU13189.  
XX  
PT Isolated nucleic acid molecules encoding musculoskeletal system  
XX associated polypeptides, useful for detecting disorders, e.g. cancer.  
XX  
XX Claim 1; SEQ ID NO 819; 321pp; English.

The invention describes an isolated nucleic acid molecule comprising a sequence encoding musculoskeletal system associated polypeptides useful for detecting disorders, e.g., cancer or cancer metastases, in animals or humans. The nucleic acid: stimulates re-vascularisation of ischaemic tissues associated with conditions such as thrombosis, arteriosclerosis, and other cardiovascular conditions; treats wounds due to injuries, burns, post-operative tissue repair, and ulcers; stimulates angiogenesis and limb regeneration; stimulates neuronal growth; can treat and prevent neuronal damage occurring in certain disorders or neurodegenerative conditions, such as, Alzheimer's disease, Parkinson's disease, and AIDS-related complex; stimulates chondrocyte growth, thus they can be used to enhance bone and periodontal regeneration and aid in tissue transports or bone grafts; prevents skin aging due to sunburn by stimulating

keratinocyte growth; prevents hair loss, since RGF family members activate hair-forming cells and promotes melanocyte growth; stimulates growth and differentiation of hematopoietic cells and bone marrow cells when used in combination with other cytokines; maintains organs before transplantation or for supporting cell culture of primary tissues; CC induces tissue of mesodermal origin to differentiate in early embryonic CC increases or decreases the differentiation or proliferation of embryonic CC stem cells, besides, haematopoietic lineage; modulates mammalian CC characteristics, such as, body height, weight, hair colour, eye colour, CC skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., CC cosmetic surgery); modulates mammalian metabolism; changes mammal's metal CC state or physical state by influencing biorhythms, circadian rhythms, CC depression, tendency for violence, tolerance for pain, reproductive CC capabilities, hormonal or endocrine levels, appetite, libido, memory, or CC stress; increases or decreases storage capabilities, fat content, lipid, CC protein, carbohydrate, vitamins, minerals, cofactors or other nutritional CC components. This sequence encodes a novel human musculoskeletal system CC antigen. Note: The sequence data for this patent did not form part of the CC printed specification, but was obtained in electronic format directly CC from the US patent office at  
XX ftp://seqdata.uspto.gov/sequence.html?DocID=20020147140

Sequence 366 BP; 120 A; 59 C; 53 G; 133 T; 0 U; 1 Other;  
SQ  
Query Match 5.1%; Score 37; DB 8; Length 366;  
Best Local Similarity 50.6%; Pred. No. 1.2;  
Matches 88; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 549 AGAGATTAGCAAGCGAATAAACTATGCTAGCTGGAGTGTGTTTCAATCATATATA 608  
DB 205 AGATTATTTTAAATGAATAATTAATTAATGCGCTTTTAAATGTTGCTTAATTTGAAA 146  
QY 609 AAGGAGAAATTTGTTCTCCTATGCTATGACAGTTTCTGGACGCTTAACTTTATTTGACG 668  
DB 145 GAAGGACCAAGTGTCTCTATGCTGCAACCAATTCAAATGTTGGCTTTTAGTAA 86  
QY 669 AGACATTCAAATCATACGATATTTGTCAAAAAAAGACTATATATA 722  
DB 85 AACTCTTCAATTAANACAGTCATTATACCGTTAAAAAATCCAGATTGATA 32

Search completed: February 7, 2005, 21:23:36  
Job time : 534 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2005, 21:13:12 ; Search time 3214 Seconds  
(without alignments)  
8562.684 Million cell updates/sec

Title: US-10-776-213-2

Perfect score: 723  
Sequence: 1 ccttcgattacgcacgcacac.....aaaagacataataacat 723

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST: \*  
1: gb\_ests1:\*  
2: gb\_ests2:\*  
3: gb\_hc2:\*  
4: gb\_ests3:\*  
5: gb\_ests4:\*  
6: gb\_ests5:\*  
7: gb\_ests6:\*  
8: gb\_ests7:\*  
9: gb\_ests8:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	249	34.4	434	8	AQ874584 V11B3 mt
C 2	140.8	19.5	927	9	AL398217 T3 end of
C 3	45	6.2	821	9	CNS06090X
C 4	42.8	5.9	1055	8	CC252581 CH261-136
C 5	42	5.8	1101	9	CNS0178X
C 6	41.6	5.8	1101	9	CNS01232
C 7	41.4	5.7	436	4	BM274028 Drosophila
C 8	41.4	5.7	1101	9	CNS0100X
C 9	40.8	5.6	791	9	AG532718 Mus muscu
C 10	40.4	5.6	698	8	BH572702 BOSTD44TF
C 11	40.4	5.6	791	4	BI255759 602971180
C 12	40.2	5.6	1000	9	CNS020UV
C 13	40	5.5	366	4	BG628729 CC-est1CL
C 14	40	5.5	518	1	AU077905 AU077905
C 15	40	5.5	845	8	AZ541311 ENTMD65TR
C 16	40	5.5	871	8	AZ674746 ENTMD28TF
C 17	39.6	5.5	179	2	AM477009 GA40E04.Y
C 18	39.6	5.5	665	6	CA126956 SCVPLR104
C 19	39.6	5.5	668	9	CE377922 tigr-g88-
C 20	39.6	5.5	677	9	CL814372 OR-GBA003
C 21	39.6	5.5	716	9	CL814356 OR-GBA003
C 22	39.6	5.5	1332	3	CR675673 Tetradon
C 23	39.4	5.4	641	9	CE491438 tigr-g88-
C 24	39.2	5.4	397	5	BQ399446 NISC_mp03

C 25	39	5.4	801	7	CK461301
C 26	39	5.4	1233	8	CC192318
C 27	38.8	5.4	932	8	AQ752307 HS-5565_B
C 28	38.6	5.3	553	7	CN385766 LB2TR04H2
C 29	38.6	5.3	630	8	B19521 T907-T7 TAM
C 30	38.6	5.3	677	9	CE498419 tigr-g88-
C 31	38.6	5.3	1081	9	CL510366 SAILL_829
C 32	38.6	5.3	1101	9	CNS00370
C 33	38.4	5.3	243	2	AM633159
C 34	38.4	5.3	594	8	AQ383493 RPI11-15
C 35	38.4	5.3	739	9	CE806390 tigr-g88-
C 36	38.2	5.3	1316	9	AG435289 Mus muscu
C 37	38.2	5.3	398	2	AM133335
C 38	38.2	5.3	674	9	AG157962 Pan troyi
C 39	38	5.3	600	8	BZ312600
C 40	38	5.3	600	8	BZ327427
C 41	38	5.3	691	9	CL951462 OeIRUA000
C 42	38	5.3	802	9	AG468867 Mus muscu
C 43	38	5.3	914	8	AZ547087 ENTMCX40TR
C 44	38	5.3	931	8	AZ679612 ENTU83TR
C 45	38	5.3	2639	3	AF289590 Homo sapi

#### ALIGNMENTS

RESULT 1  
AQ874584/c  
LOCUS  
DEFINITION  
V11B3 mtH-3xHA/lacZ insertion library, strain Y2278 Saccharomyces  
cerevisiae genomic 5', genomic survey sequence.

ACCESSION  
AQ874584  
VERSION  
AQ874584.1 GI:6286828  
KEYWORDS  
GSS.  
SOURCE  
Saccharomyces cerevisiae (baker's yeast)  
ORGANISM  
Saccharomyces cerevisiae

#### REFERENCE

1 (bases 1 to 434)  
Rosa-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,  
desRages, S.A., Cheung, K.-H., Sheehan, A., Symonatsis, D., Jansen, R.,  
Umanak, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,  
Hager, K., Miller, P., Roeder, G.S. and Snyder, M.  
Large-Scale Analysis of the Yeast Genome by Transposon Tagging and  
Gene Disruption  
Unpublished (1999)

#### JOURNAL COMMENT

Contact: Kumar A  
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology  
Yale University  
P.O. Box 208103, New Haven, CT 06520-8103, USA  
Tel: 203 432 9949  
Fax: 203 432 6161  
Email: anuj.kumar@yale.edu  
te of mtH-3xHA/lacZ insertion.  
Seq primer: GGCCTTCTTCTTGAAGATAC  
Class: transposon-tagged.  
Location/Qualifiers  
1. 434

#### FEATURES

source

/organism="Saccharomyces cerevisiae"  
/mol\_type="genomic DNA"  
/strain="Y2278 - 5288C background, cir(0) rho(0)"  
/db\_xref="taxon:4932"  
/lab\_host="E. coli"  
/clone\_lib="mtH-3xHA/lacZ insertion library, strain Y2278"  
/note="Vector: pHS6-Sal; A yeast genomic DNA library  
without 2 micron or mitochondrial DNA was prepared in  
pHS6-Sal; genomic DNA was size-fractionated (DNA of  
roughly 2-3 kb in length) prior to cloning. This library  
was subsequently mutagenized with a mtH-3xHA/lacZ  
multitransposon containing lacZ, URA3, and tet resistance."

#### ORIGIN

Query Match	34.4%	Score 249	DB 8	Length 434
Best Local Similarity	100.0%	Pred. No. 6,2e-60		
Matches 249	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	1	CTTTCGATTGACGACGACACACATCATGATCGTGCATATAAATACTACGAGAAA	60	
Db	311	CTTTCGATTGACGACGACACACATCATGATCGTGCATATAAATACTACGAGAAA	252	
Qy	61	ACCATTAAGAGCAAGAACGATACCTACTTGGAAAGAAAGAGACGCTGTGAAGGGGAT	120	
Db	251	ACCATTAAGAGCAAGAACGATACCTACTTGGAAAGAAAGAGACGCTGTGAAGGGGAT	192	
Qy	121	GGGGGCTAAGAAATCATTCATCTTTTCCTTCGCGGTCCGAGCCCGAGACCCCTCCT	180	
Db	191	GGGGGCTAAGAAATCATTCATCTTTTCCTTCGCGGTCCGAGCCCGAGACCCCTCCT	132	
Qy	181	CTCCCCGACGAGATTTCTTCTTTCATATCTTCTTTATTTCTTATCCGTTAGACACC	240	
Db	131	CTCCCCGACGAGATTTCTTCTTTCATATCTTCTTTATTTCTTATCCGTTAGACACC	72	
Qy	241	GCACATATGA 249		
Db	71	GCACATATGA 63		
RESULT 2				
CNS06GVN		927 bp	DNA	linear
LOCUS		T3 end of clone AS0AA007G06 of library AS0A from strain CLB 533		
DEFINITION		T3 end of clone AS0AA007G06 of library AS0A from strain CLB 533		
ACCESSION		AL398217		
VERSION		AL398217.1		
KEYWORDS		GS.		
SOURCE		Saccharomyces bayanus		
ORGANISM		Saccharomyces bayanus		
REFERENCE		Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
AUTHORS		1 (bases 1 to 927)		
		Soucier, J. L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brothier, P., Casaregola, S., de-Montigny, D., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, M., Tekala, F., Toffano-Nicche, C., Wesolowski-Louvel, M., Wincker, P. and Weissenbach, J.		
		Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies		
		FEBS Lett. 487 (1), 3-12 (2000)		
		20584711		
		11152876		
		2 (bases 1 to 927)		
		Bon, E., Neuveglise, C., Casaregola, S., Artiguenave, F., Wincker, P., Aigle, M. and Durrens, P.		
		Genomic exploration of the hemiascomycetous yeasts: 5. Saccharomyces bayanus var. uvatum		
		FEBS Lett. 487 (1), 37-41 (2000)		
		20584715		
		11152880		
		3 (bases 1 to 927)		
		Genoscope.		
		Direct Submission		
		Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaillon Creteil, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)		
		This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvatum, Saccharomyces exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.		
COMMENT				

FEATURES	Location/Qualifiers
source	1..927
	/organism="Saccharomyces bayanus"
	/mol_type="genomic DNA"
	/strain="CLIB 533"
	/variety="uvatum"
	/db_xref="taxon:4931"
	/clone="AS0AA007006"
	/clone_1ib="AS0AA"
	/note="end : 73"
	<3..>107
	/note="similar to Saccharomyces cerevisiae ORF YMR251w [
	strong similarity to YKR076w and YGR154c ]"
	/evidence=not_experimental
ORIGIN	
Query Match	19.5%; Score 140.8; DB 9; Length 927;
Best Local Similarity	69.2%; Pred. No. 6.7e-29;
Matches 321; Conservative	0; Mismatches 107; Indels 36; Gaps 8
QY	114 GGGGATGGGGCTTAAAGATCATCTTCTTTCCCTTCGGGGTCCGACCCGGAC 173
DB	240 GCGAAGAGGGGGGTGGAAGTCACTCAC-TCTCTTCCCTTTATAGTCCGACCCGAAA 298
QY	174 CCTCTCTCTCCCGACGATTTCTTCTTCAATCTTCTTTATTCCTATCCCGTTGA 233
DB	239 CCCCCCTTCCCGACGCGTTTCTTGTTCT------TCTTTTCCCGTTGG 345
QY	234 AGCAACGCGACATATGACATTAATATGATGCGACATCTCATAGCTGTGTGTATC 293
DB	346 GACACACGACATGAGCATTAAGAGACGCTGGCTTTCTATATGCTGAT-----TGTCCCA 399
QY	294 TCACAGTGTATACGCGACCGTGGCTCGGAAACGTTCTCTGTGACATTTCTAGAACAG 353
DB	400 AAACGACATATGCGACATCGGGCCCGAAC--GTCTCTTAATAGTCTTGAAGCG 457
QY	354 GGTTCAGTCTCGATTAAT-AGAAATATAGCGCATTTTGTAGCGCCGCCGCG- 406
DB	458 GGTATATGCGCTTGACAAATGCTTAATATACGCGCATTTTGTAGCGCGCGGTGTGCTG 517
QY	407 -----GCGCCCGTGTCCCAATAGGAGGCGGAGTTATCGCGGAGAGCTTACTCTTCC 461
DB	518 CGCAACGCGCCCTTCTCTAAATAGGAGGCGGAGTTATCGCACTGCTTAC-TCTTCT 576
QY	462 ATTGGGTAAACCCCTTCTTGTCTTTTGGCGCAGTGTGTGACAGTCTGCGCGGAGAAC 521
DB	577 ATTGGGGGAAAGCCCTTCTTGTCTTTTCCGCCAGAGGCTGACAGAGCTGCGCGGAGATA 636
QY	522 TAGATATAGGAGATGTATAC--TTTCGATGAGAGATTTAGCAAGCG 564
DB	637 GAGCGATTAAGGATGTACTTTTTCATGAGAGACTTACAGAGAG 680
RESULT 3	
LOCUS	CNS0090X 821 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence JET3 end of BAC #
VERSION	BAK19021 of RFL-98 library from Drosophila melanogaster (fruit
KEYWORDS	fly), genomic survey sequence.
ACCESSION	AL052985
VERSION	AL052985.1 GI:4934433
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster (fruit fly)
ORGANISM	Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
AUTHORS	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
TITLE	Ephydroidea; Drosophilidae; Drosophila.
JOURNAL	1 (bases 1 to 821)
	Genoscope.
	Direct Submission
	Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage :
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
	- Web : www.genoscope.cns.fr)

COMMENT	FEATURES	ORIGIN
Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoler in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .	<p>1..821</p> <p>location/Qualifiers</p> <p>organism="Drosophila melanogaster"</p> <p>molecule_type="Genomic DNA"</p> <p>db_xref="taxon:7227"</p> <p>clone="BACR19021"</p> <p>clone_library="RPc1-98"</p> <p>note="end : TEf3"</p>	
Query Match Best Local Similarity 37.9%; Pred.No.0.18; Matches 99; Conservative 38; Mismatches 124; Indels 0; Gaps 0;		
Qy 462 ATTGGGGAAGCCCTTCTCTGTTTGGCCAGTGTCTGAGCTCGCCGAGAAACA 521		
Db 561 AAYKDKACAGTGGCTCTTTTWTWKTCGCCBATAATAMCCCRATGCGCTTTTTCMCM 502		
Qy 522 TAGTGATPAAGGAATGTAACTTTCAGAGAGAAATAGCAAGCGAAAAAATAATGCTTA 581		
Db 501 WAMATATTCACACTATMTWTTTBTBXTATTAADACAMGASAMWATOCATAST 442		
Qy 582 GCTGGAGACTTTTTCATCATATAAAGGAGAAATGTTGCTCACTATGTGACAGTT 641		
Db 441 WTGAATPAAGAGTTTATATGCTTCTCKAAAMCTGAAATACATWTCCKMKKCGGC 382		
Qy 642 TCTGGAGAGCTTAATCTTTATATGAGAGAGCTATCAATCATATACAGATATGTCAAAA 701		
Db 381 TACGCAATCTATATATWTTTWTATAAAAAABATATATATATTAATTAATTAATAA 322		
Qy 702 AAAAAAGACTAATATATACA 722		
Db 321 WAAAAATPAAAAAWMAATA 301		
RESULT 4 CC252581/c		
LOCUS CH261.136h21.Sp6.1 CH261 Gallus gallus genomic clone CH261-136h21, genomic survey sequence.		
DEFINITION CC252581		
ACCESSION CC252581		
VERSION CC252581.1 GI:30589331		
KEYWORDS GSS.		
SOURCE Gallus gallus (chicken)		
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Gallus; Phasianinae; Gallus.		
REFERENCE 1 (bases 1 to 1055) Kremetzki,C., Higinbotham,J., Wylie,K., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R. Gallus gallus BAC End Reads Unpublished (2003) Contact: Richard K. Wilson Genome Sequencing Center Washington University School of Medicine Email: <a href="mailto:submissions@wustl.edu">submissions@wustl.edu</a> Insert Length: 182000 StdError: 0.00		
TITLE		
JOURNAL		
COMMENT		

```

Seq primer: Sp6 ATTAGTCACACTATAG
Class: BAC ends
High quality sequence start: 177
High quality sequence stop: 279.
Location/Qualifiers
1. 1055
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-136H21"
/sex="Female"
/cell_line="UCD001, inbred 256"
/clone_1lb="CH261"
/name="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - For library and clone
ordering information: http://www.chori.org/bacpac"

ORIGIN
Query Match 5.9%; Score 42.8; DB 8; Length 1055;
Best Local Similarity 55.3%; Pred. No. 0.79;
Matches 83; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

Oy 472 GCCCCTTTCGTGTTCCGCCACAGTGTCTCTGACGCTGCCCGGAGAACATAGCATTAAG 531
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 674 GTCCCCCTCTATTTTCATATAGTCTCGAATTCCTGCAGACTGAGATTAATAGGTGTGA 615
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 532 GGATGTAACTTTCGATGAGAAATTAGCAAGCGAATAAATATAGCTCTAGCTGGAGTT 591
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 614 GAGACCAACCTGTAGAGGAGAAATATATGAGAGCGAATAATATGATGAGAGGAGTT 555
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 592 GTTTTCAATCATATTAAGGAGAAATTG 621
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 554 CTTTCCAATATATAGAGAGAGAGATGTTG 525
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
CNS017KX/c 1101 bp DNA linear GSS 26-JUL-1999
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence Sp6 end of BAC
BACN37F10 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
A1108171
A1108171.1 GI:5628475
GSS.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (base 1 to 1101)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (edgp) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelosAC11.
Location/Qualifiers
1. 1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_1lb="BACN37F10"
/clone_1lb="DrosBAC"
/plasmid="pBelosAC11"
/note="end : Sp6"

```



AG532718	AG532718	791 bp	DNA	linear	GSS 05-JUN-2004
LOCUS	Mus musculus molossinus DNA, clone:MSMg01-443B08.TU, genomic survey				
DEFINITION	sequence.				
ACCESSION	AG532718				
VERSION	AG532718.1	GI:48293132			
KEYWORDS	GSS.				
SOURCE	Mus musculus molossinus				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1				
AUTHORS	Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.				
TITLE	BAC end Sequences of Library MSMg01				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 791)				
AUTHORS	Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)				
COMMENT	Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp) Tsukuba Institute, Bio Resource Center, The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan phone: 81-298-36-9189, fax: 81-298-36-9199				
	e-mail: abe@rtc.riken.jp				
	PRIMERS				
	Sequencing : TU				
	LIBRARY				
	Vector : pBACe3.6				
	R.Site 1 : EcoRI				
	R.Site 2 : EcoRI.				
FEATURES	Location/Qualifiers				
Source	1..791				
	/organism="Mus musculus molossinus"				
	/mol_type="genomic DNA"				
	/sub_species="molossinus"				
	/db_xref="taxon:57486"				
	/clone="MSMg01-443B08.TU"				
	/sex="male"				
	/tissue_type="mixture of kidney and spleen"				
	/clone_lib="MSMg01 Mouse Male BAC library"				
ORIGIN					
Query Match	5.6%; Score 40.8; DB 9; Length 791;				
Best Local Similarity	55.6%; Pred.No.2.8;				
Matches	75; Conservative 0; Mismatches 60; Indels 0; Gaps 0;				
Y	110 GTAAGGGGAGATGGGGGCTAAGAAGTCATTCCTTTCCCTTCGCGATCCGAGCCG 169				
Db	290 GAAAGCGGAGATGGGGGAGCGTTTTCCTTTCCNCCCCCCCCCNCNCCCCC 349				
Y	170 GAAACCCCTCCTCTCCCGCAGATTTCTTCCTTCATATCTTCCTTTATTCATCCG 229				
Db	350 CCCCCCTTTTCCCTCCCTTTTCTCTCTGCTTCCCTTCTCTTTTTTTC 409				
Y	230 TTGAAGCAACCGCAC 244				
Db	410 TTTTTCACCTCC 424				
RESULT 10					
BH572702	BH572702	698 bp	DNA	linear	GSS 14-DEC-2001
LOCUS	BOGTD44TF BOGT Brasica oleracea genomic clone BOGTD44, genomic				
DEFINITION	survey sequence.				
ACCESSION	BH572702				
VERSION	BH572702.1	GI:17824541			
KEYWORDS	GSS.				

SOURCE	Brassica oleracea
ORGANISM	Brassica oleracea
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
AUTHORS	1 (bases 1 to 698)
TITLE	Towm,C.D., Van Aken,S., Uiterback,T., Koo,H. and Fraser,C.M. Whole genome shotgun sequencing of Brassica oleracea Unpublished (2001)
JOURNAL	Other_GSSs: BCGTMD4TR
COMMENT	Contact: Chris Town TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA. Tel: 301-838-3523 Fax: 301-838-0208 Email: cdromw@ig.org DNA is from a doubled haploid provided by Tom Osborn. Seq primer: 1P Class: sheared ends.
FEATURES	Location/Qualifiers
source	1..698 /organism="Brassica oleracea" /mol_type="genomic DNA" /strain="TO100DH3" /db_xref="taxon:3712" /clone_1fb="BOGTD44" /clone_1lb="BOGT" /note="Vector: PHOS1, Site 1; BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"
ORIGIN	
Query Match	5.6%; Score 40.4; DB 8; Length 698;
Best Local Similarity	50.0%; Pred. No. 3.5;
Matches	101; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
OY	518 AACATAGCATGATGAGGATGACTCTTCGATGGAGAGAATTAGCAACGCCGAATAAACTAGT 577
DB	485 AAAAATTTCTTACTCAATTCGTTTTTGCGAGGATTTAGTACTGATGATTAATAATTTGTTA 544
OY	578 GCTAGCTGGAGCTTGTCTTTCATCATATAAAAGGAGAAATTTGCTCCTACTATGTGAC 637
DB	545 ACTAAGTTAGATATATTTTGCACATATATTAGATATTTTAGTTTGGGTTTAATAATATA 604
OY	638 AGTTCTGGGACGCTTAACTTTTATTCGACAGGACTATCAAATATATACAGATATTGCA 697
DB	605 AATTAGTTTGATGATATATCCAAATTATGATTTAAAAAACAATTTTTTTA 664
OY	698 AAAAAAAAAAGACTAATATA 719
DB	665 AAAAAATACCATTAATATATA 686
RESULT 11	
B1255759/c	
LOCUS	B1255759 791 bp mRNA linear EST 17-JUL-2001
ACCSSION	6029771180F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5122410 5',
VERSION	mRNA sequence.
KEYWORDS	B1255759
SOURCE	B1255759.1 GI:14809497
ORGANISM	EST.
Homo sapiens (human)	
Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
NIH-MGC http://mgc.nci.nih.gov/.	
Unpublished (1999)	
National Institutes of Health, Mammalian Gene Collection (MGC)	
Contact: Robert Strausberg, Ph.D.	
Email: cgabbs-remail.nih.gov	
Tissue Procurement: ATCC	
cDNA Library Preparation: Life Technologies, Inc.	
cDNA Library Arrayed by: Incyte Genomics, Inc.	

FEATURES	source
<p>DNA sequencing by: Incyte Genomics, Inc.  Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/MLN at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>  Plate: LHAM11298 row: 0 column: 19  High quality sequence start: 2  High quality sequence stop: 711.  Location/Qualifiers  1..791</p>	<p>/organism="Homo sapiens"  /mol_type="mRNA"  /db_xref="taxon:9606"  /clone="IMAGE:5122410"  /issue_type="cervical carcinoma cell line"  /lab_host="DH10B"  /clone_id="NH MGC 12"  /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: Ncrl; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  Average insert size 1.4 kb. Library prepared by Life Technologies."</p>
ORIGIN	
Query Match	5.6%; Score 40.4; DB 4; Length 791;
Best Local Similarity	58.2%; Pred. No. 3.6; Mismatches 51; Indels 0; Gaps 0;
Matches	71; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
Qy	588 AGTGTGTTTCATCATATATAAAGGAGAAATTGTCTCACTATGACAGATTCTGGG 647
Db	505 ACTCTCTCACACATCAATAGTCAGTGAANTTTACAGTGGGANTTCTACAAATCAGG 446
Qy	648 ACGCTTAACTTTATTATGACAGAGCATCTCAATCATACAGATATGTCTCAAAAAA 707
Db	445 ATTCTTAAATTTTGTGTGACGATTATCTGAGAGTATAAGTTTAAAGAAAAA 386
Qy	708 AG 709
Db	385 TG 384
RESULT 12	
CNS02UHV	1000 bp DNA linear GSS 01-SEP-2000
LOCUS	Tetradon nigriviridis genome survey sequence T7 end of clone
DEFINITION	167C22 of library G from Tetradon nigriviridis, genomic survey sequence.
ACCESSION	AL214492.1 GI:7873311
VERSION	GSS; genome survey sequence.
KEYWORDS	Tetradon nigriviridis
SOURCE	Tetradon nigriviridis
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetradon.
REFERENCE	
AUTHORS	1 Roest Crolius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizeser,C., Winckler,P., Broctier,P., Quetier,F., Saurin,W. and Weissenbach,J. Estimate of human gene number provided by genome-wide analysis using Tetradon nigriviridis DNA sequence Nat. Genet. 25 (2), 235-238 (2000)
JOURNAL	20296633
MEDLINE	10835645
REFERENCE	
AUTHORS	2 Roest Crolius,H., Jalllon,O., Dasilva,C., Ozouf-Costaz,C., Fizeser,C., Fisches,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigriviridis
JOURNAL	Genome Res. 10 (7), 939-949 (2000)
MEDLINE	20359837
PUBMED	10899143
REFERENCE	3 (bases 1 to 1000)



**AUTHORS** Genoscope.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : segre@genoscope.cns.fr)  
**COMMENT** Web : www.genoscope.cns.fr  
 This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetradon>.  
**FEATURES**  
 source  
 1..1000  
 /organism="Tetradon nigroviridis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:99883"  
 /clone\_1="167C22"  
 /clone\_11b="G"  
 /note="Genoscope sequence ID : COAG167B11LPL-end : T"  
**ORIGIN**  
 Query Match 5.6%; Score 40.2; DB 9; Length 1000;  
 Best Local Similarity 45.4%; Pred.No.4.3;  
 Matches 59; Conservative 19; Mismatches 52; Indels 0; Gaps 0;  
**Qy** 91 AAGGAAAAAGGACGCTTTAAGGGGGAGTGGGGGCTAAGAAGCATTCACCTTCTTTC 150  
 |||||  
 851 AAAAAAAGGGGGGKGGGGGGGGGGGGGGGGGGGGGAGANANKKTKKKKBYYYYYYY 910  
 ::|||  
**Db** 911 YYCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTTTTTTTTTTTTTTTTTT 970  
 |||||  
**Qy** 211 TCCTTTATT 220  
 |||||  
**Db** 971 TTTT TTTT 980  
**RESULT 13**  
**LOCUS** BG628729/c 366 bp mRNA linear EST 19-APR-2000  
**DEFINITION** CC-ef1cEL23G13d1 Tomato flower library from a mixture of developmental stages Lycopersicon esculentum cDNA clone  
**ACCESSION** BG628729 GI:13680202  
**VERSION** BG628729.1 GI:13680202  
**KEYWORDS** EST.  
**SOURCE** Lycopersicon esculentum (tomato)  
**ORGANISM** Lycopersicon esculentum  
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; Lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.  
**REFERENCE** 1 (bases 1 to 366)  
**AUTHORS** van der Hoeven,R.S. and Tanksley,S.D.  
**TITLE** ESTs from a tomato flower library  
**JOURNAL** Unpublished (2001)  
**COMMENT** Contact: Rieger S. van der Hoeven  
 Cornell University  
 252 Emerson Hall, Ithaca, NY 14850, USA  
 Tel: 607 255 7886  
 Fax: 607 255 6683  
 Email: rv19@cornell.edu  
 3 prime sequence.  
**FEATURES**  
 source  
 1..366  
 Location/Qualifiers  
 /organism="Lycopersicon esculentum"  
 /mol\_type="mRNA"  
 /cultiVar="B6203"  
 /db\_xref="taxon:4081"  
 /clone="CC-ef1cEL23G13d1"  
 /tissue\_type="developing flower buds and open flowers"  
 /dev\_stage="4-8 week old plants"  
 /lab\_host="X10LR"  
 /clone\_lib="Tomato flower library from a mixture of developmental stages"

ORIGIN	Query Match	5.5%; Score 40; DB 4; Length 366;	Best Local Similarity 55.9%; Pred. No. 4.1;	Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
OY	584	TGGAGTTGTTTTCATATCAATAAAGGAGAAATGTGCTCATATGTGACAGTTTC	643	
Db	136	TGGAAGTTGATTGATTAATTTAGATTAAGATTAAGTTTCTTTGGGCTAATATGTGT	77	
OY	644	TGGACGCTTAATCTTTATTTGACAGAGACATATCAATCATACGATATTTGTCAAAAAA	703	
Db	76	CAGACCTATTTTCTTTTGGCTTATTTTGTATTCAGATGAAATATTTTCTAAAAAAA	17	
OY	704	AAAAAGACTAATAATA 719		
Db	16	AAAAAAAAAAAAAAAA 1		
RESULT 14				
LOCUS	AU077905	518 bp	mRNA	linear
DEFINITION	AU077905	Rice shoot	Oryza sativa (japonica cultivar-group)	CDNA
ACCESSION	AU077905			
VERSION	AU077905.1	GI:5667645		
KEYWORDS	EST.			
SOURCE	Oryza sativa (japonica cultivar-group)			
ORGANISM	Oryza sativa (japonica cultivar-group)			
REFERENCE	Yamamoto, K. and Sasaki, T.			
AUTHORS	Rice cDNA from etiolated shoot			
TITLE	Unpublished (1997)			
JOURNAL	Contact: Takuji Sasaki			
COMMENT	National Institute of Agrobiological Resources			
	Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki			
	305-8602, Japan			
	Tel: 81-298-38-7441			
	Fax: 81-298-38-7468			
	Email: tassaki@abr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/			
	PROJECT = 'RGP'.			
FEATURES	Location/Qualifiers			
SOURCE	1..518			
	/organism="Oryza sativa (japonica cultivar-group)"			
	/mol_type="mRNA"			
	/cultivar="Nipponbare"			
	/db_xref="taxon:39947"			
	/clone="SS033.8Z"			
	/dev_stage="Etiolated shoot (8 days old)"			
	/clone_id="Rice shoot"			
	/note="Etiolated shoot (8 days old)"			
ORIGIN				
Query Match	5.5%; Score 40; DB 1; Length 518;			
Best Local Similarity	57.0%; Pred. No. 4.3;			
Matches	73; Conservative 0; Mismatches 55; Indels 0; Gaps 0;			
OY	595	TTTCAATCATTAAGAAGAAATTTGTTGCTCATATGACAGATTTCGGAGCTCTT	654	
Db	388	TCTGAATTTGTTAGTAGGAAGAAGATTGTTATTCGTGGTGATCATCAAAATGCTGT	447	
OY	655	AACCTTTATTCGACAGACATATCAATCATACAGATATTTGTCAAAAAAAGACTAA	714	
Db	448	AATTTGCTCTCAGATGACATGGAATCATATCAGAGTTGATATRAAAAAAAAAAAAAA	507	
OY	715	TAAATACA 722		

Db 508 AAAAAAA 515

RESULT 15  
 AZ541311 845 bp DNA linear GSS 14-NOV-2000  
 LOCUS ENTDV657R Entamoeba histolytica Sheared DNA Entamoeba histolytica  
 DEFINITION genomic, genomic survey sequence.  
 ACCESSION AZ541311 GI:11148922  
 VERSION AZ541311.1 GI:11148922  
 KEYWORDS GSS.  
 SOURCE Entamoeba histolytica  
 ORGANISM Entamoeba histolytica  
 Eukaryota; Entamoebidae; Entamoeba.  
 REFERENCE 1 (bases 1 to 845)  
 Authors Loftus, B., Van Aken, S. and Fraser, C.  
 TITLE Determination of clone end sequences from Entamoeba histolytica  
 HMI:IMSS sheared DNA library  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: b.loftus@tigr.org  
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
 DNA library  
 Seq primer: M13-Reverse  
 Class: Shotgun  
 High quality sequence start: 42  
 High quality sequence stop: 612.  
 Location/Qualifiers

# FEATURES

1..845  
 /organism="Entamoeba histolytica"  
 /mol\_type="genomic DNA"  
 /strain="HMI:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_id="Entamoeba histolytica Sheared DNA"  
 /note="Vector: pHOSt; Site 1; Bst I; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from broth cultures of E. histolytica  
 using a method described by Clark and Diamond (Clark,  
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
 method for isolate identification. Exp. Parasitol.  
 77:450.). The DNA was mechanically sheared to give a  
 tight size distribution (~2 kb). The v + i method used for  
 the library construction is described in detail in Smith,  
 H.O. and Venter, J.C. (Making small insert libraries for  
 whole genome shotgun sequencing projects. in Genome  
 Sequencing: A Practical Approach, eds. W. Vaudin and B.  
 Barrell, Oxford University Press, 1999)."

## ORIGIN

Query Match 5.5%; Score 40; DB 8; Length 845;  
 Best Local Similarity 55.9%; Pred.No. 4.8; Indels 0; Gaps 0;  
 Matches 76; Conservative 0; Mismatches 60;  
 Oy 588 AGTGTCTTTCATCATATAAGGAGAAATGTTGCTCACTATGTGACAGTTCTGG 647  
 Db 407 AATTATTTAAATAAATATCATTTTGGATATTTTATTGTTAATTGACAGAACTGAA 466  
 Oy 648 ACGTCTTACTTTTATTCAGAGACTATCAATCATACAGATATGTCAAAAAAAA 707  
 Db 467 AGTTCTTAATTTAATAGTAAGAAATATTAATCAATGATGCTTTTCTTAACAAAAAT 526  
 Oy 708 AGACTAATATACAT 723  
 Db 527 TTATGAAAAATATAT 542

Search completed: February 7, 2005, 23:17:23  
 Job time : 3219 secs

QY 1 CTTTCGATTGACGCGACACACATCATCATAGACCTGCGCTCTAAATAATACCTACCGGAAAA  
Db 1 CTTTCGATTGACGCGACACACATCATCATAGACCTGCGCTCTAAATAATACCTACCGGAAAA  
QY 61 ACCATAAAGGCAAAAGCGATACCTGCTTGGAGAAAAAGAGCAGCGTTGTAAAGGGGGA  
Db 61 ACCATAAAGGCAAAAGCGATACCTGCTTGGAGAAAAAGAGCAGCGTTGTAAAGGGGGA  
QY 121 GGGGGCTAAGAAGCATTCCTTTCTTTGCGGCGTCCGGAACCGGGAGCCCTCTC  
Db 121 GGGGGCTAAGAAGCATTCCTTTCTTTGCGGCGTCCGGAACCGGGAGCCCTCTC  
QY 181 CTCGCCGACGATTCTTCCCTTTCATATCTTCTTTATTCCTATCCCGTTGAAGCAAC  
Db 181 CTCGCCGACGATTCTTCCCTTTCATATCTTCTTTATTCCTATCCCGTTGAAGCAAC  
QY 241 GCACTATGACATAATGSGCTGGGCACTCCATGSGCGTGAATGATGATCAAGT  
Db 241 GCACTATGACATAATGSGCTGGGCACTCCATGSGCGTGAATGATGATCAAGT  
QY 301 GGTAAAGGCAACCGTGCTCGAAAAAGGTTCTTCGTGACAATTCAGAACAGGGGCTACG  
Db 301 GGTAAAGGCAACCGTGCTCGAAAAAGGTTCTTCGTGACAATTCAGAACAGGGGCTACG  
QY 361 GTCTCATATATGAATATATAGGCAATTTTGTCTAGGCCGCGCGGCGCCGTTTCCG

```
Db      361  GTCGATATATAGATATATTAAGCGCATTTTTCGTAGCCGCCGCCGCCGCTTTTCCCA 420
Qy      421  ATAGGAGAGCGCAGTTTATTCGGCGAGCTCTAATTTCTTATTTGGTAAAGCCCTTTC 480
Db      421  ATAGGAGAGCGCAGTTTATTCGGCGAGCTCTAATTTCTTATTTGGTAAAGCCCTTTC 480
Qy      481  TGTTCGCGCAGTGTGCTGCGAGGCTGCCCGGAGAACATATGTAATTAAGGATGTAAAC 540
Db      481  TGTTCGCGCAGTGTGCTGCGAGGCTGCCCGGAGAACATATGTAATTAAGGATGTAAAC 540
Qy      541  TTTTCGATGAGAAATTATTCAGAACGCGGAAAAAACTATGCTAGCTGGAGTGTTCCTAA 600
Db      541  TTTTCGATGAGAAATTATTCAGAACGCGGAAAAAACTATGCTAGCTGGAGTGTTCCTAA 600
Qy      601  TCATATTAATAAGGAGAAATTTGTGCTCATCTATGTGACAGTTTCTGGGACGTCTTAACCTT 660
Db      601  TCATATTAATAAGGAGAAATTTGTGCTCATCTATGTGACAGTTTCTGGGACGTCTTAACCTT 660
Qy      661  TATTGCAAGAGACTATCAATCATACAGATATTTGTCAAAAAAATAAGTAAATATATAA 720
Db      661  TATTGCAAGAGACTATCAATCATACAGATATTTGTCAAAAAAATAAGTAAATATATAA 720
Qy      721  CAT 723
Db      721  CAT 723
```

## RESULT 2

```
US-09-743-194-24
; Sequence 24, Application US/09743194
; Patent No. 6716601
; GENERAL INFORMATION:
; APPLICANT: Belfield, Graham
; APPLICANT: Oakley, Caroline
; TITLE OF INVENTION: Compositions and Methods Utilizing Sequences for
; FILE REFERENCE: 3526, 82543
; CURRENT APPLICATION NUMBER: US/09/743,194
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 11427
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-743-194-24
```

```
Query Match      100.0%; Score 723; DB 4; Length 11427;
Best Local Similarity 100.0%; Pred. No. 9, 8e-229;
Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1  CTTTCGATTAGCAGCAGACATCATATGACTGCGTCTATAAAATATACCTACGGAAAA 60
Db      15  CTTTCGATTAGCAGCAGACATCATATGACTGCGTCTATAAAATATACCTACGGAAAA 74
Qy      61  ACCATTAAGAGCAAGCGATACCTACTTGGAGAAAAAGAGACGCTTGTAAAGGGGAT 120
Db      75  ACCATTAAGAGCAAGCGATACCTACTTGGAGAAAAAGAGACGCTTGTAAAGGGGAT 134
Qy      121  GGGGGCTAAGAAATCACTTCTTTCCCTTCGCGGTCCGGACCCGGACCCCTCCT 180
Db      135  GGGGGCTAAGAAATCACTTCTTTCCCTTCGCGGTCCGGACCCGGACCCCTCCT 194
Qy      181  CTCGCCGACAGATTTCTTCCTTTCATATCTTCTTTATTTCCATCCGTTGAAGCAAC 240
Db      195  CTCGCCGACAGATTTCTTCCTTTCATATCTTCTTTATTTCCATCCGTTGAAGCAAC 254
Qy      241  GCACTATGACTAAATGCTGTGACATCTTCATGCTGTGACTGTGTATCTCAAGT 300
Db      255  GCACTATGACTAAATGCTGTGACATCTTCATGCTGTGACTGTGTATCTCAAGT 314
Qy      301  GGTAAAGGACGCTGCTCGAAAAGGTTCTTCGTGAATTTCTAGAACAGGGGCTACA 360
Db      |||
```

```
Db      315  GGTAAAGGACGCTGCTCGAAAAGGTTCTTCGTGAATTTCTAGAACAGGGGCTACA 374
Qy      361  GTCGATTAATATAGATATTAAGCGCATTTTTCGTAGCGCGCGCGCGCGCTTTTCCCA 420
Db      375  GTCGATTAATATAGATATTAAGCGCATTTTTCGTAGCGCGCGCGCGCGCTTTTCCCA 434
Qy      421  ATAGGAGAGCGCAGTTTATTCGGCGAGCTCTAATTTCTTATTTGGTAAAGCCCTTTC 480
Db      435  ATAGGAGAGCGCAGTTTATTCGGCGAGCTCTAATTTCTTATTTGGTAAAGCCCTTTC 494
Qy      481  TGTTCGCGCAGTGTGCTGCGAGGCTGCCCGGAGAACATATGTAATTAAGGATGTAAAC 540
Db      495  TGTTCGCGCAGTGTGCTGCGAGGCTGCCCGGAGAACATATGTAATTAAGGATGTAAAC 554
Qy      541  TTTTCGATGAGAAATTATTCAGAACGCGGAAAAAACTATGCTAGCTGGAGTGTTCCTAA 600
Db      555  TTTTCGATGAGAAATTATTCAGAACGCGGAAAAAACTATGCTAGCTGGAGTGTTCCTAA 614
Qy      601  TCATATTAATAAGGAGAAATTTGTGCTCATCTATGTGACAGTTTCTGGGACGTCTTAACCTT 660
Db      615  TCATATTAATAAGGAGAAATTTGTGCTCATCTATGTGACAGTTTCTGGGACGTCTTAACCTT 674
Qy      661  TATTGCAAGAGACTATCAATCATACAGATATTTGTCAAAAAAATAAGTAAATATATAA 720
Db      675  TATTGCAAGAGACTATCAATCATACAGATATTTGTCAAAAAAATAAGTAAATATATAA 734
Qy      721  CAT 723
Db      735  CAT 737
```

## RESULT 3

```
US-09-743-194-20
; Sequence 20, Application US/09743194
; Patent No. 6716601
; GENERAL INFORMATION:
; APPLICANT: Belfield, Graham
; APPLICANT: Oakley, Caroline
; TITLE OF INVENTION: Compositions and Methods Utilizing Sequences for
; FILE REFERENCE: 3526, 82543
; CURRENT APPLICATION NUMBER: US/09/743,194
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 13073
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-743-194-20
```

```
Query Match      100.0%; Score 723; DB 4; Length 13073;
Best Local Similarity 100.0%; Pred. No. 1, 1e-228;
Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1  CTTTCGATTAGCAGCAGACATCATATGACTGCGTCTATAAAATATACCTACGGAAAA 60
Db      16  CTTTCGATTAGCAGCAGACATCATATGACTGCGTCTATAAAATATACCTACGGAAAA 75
Qy      61  ACCATTAAGAGCAAGCGATACCTACTTGGAGAAAAAGAGACGCTTGTAAAGGGGAT 120
Db      76  ACCATTAAGAGCAAGCGATACCTACTTGGAGAAAAAGAGACGCTTGTAAAGGGGAT 135
Qy      121  GGGGGCTAAGAAATCACTTCTTTCCCTTCGCGGTCCGGACCCGGACCCCTCCT 180
Db      136  GGGGGCTAAGAAATCACTTCTTTCCCTTCGCGGTCCGGACCCGGACCCCTCCT 195
Qy      181  CTCGCCGACAGATTTCTTCCTTTCATATCTTCTTTATTTCCATCCGTTGAAGCAAC 240
Db      196  CTCGCCGACAGATTTCTTCCTTTCATATCTTCTTTATTTCCATCCGTTGAAGCAAC 255
Qy      241  GCACTATGACTAAATGCTGTGACATCTTCATGCTGTGACTGTGTATCTCAAGT 300
Db      |||
```

Db 256 GCATATGACTAATGCTGTGACATCTTCATGCTGTGACTTGTGTATCTCACAGT 315  
 QY 301 GGTAACGGACCGCTGCTCGGAAACGGTTCCTTCTGTGACAAATCTAGAACAGGGGCTAC 360  
 Db 316 GGTAACGGACCGCTGCTCGGAAACGGTTCCTTCTGTGACAAATCTAGAACAGGGGCTAC 375  
 QY 361 GTCCTCATTAATAGAAATTAAGGCAATTTTGTGACGCGCGCGCGCGCTTTCCCA 420  
 Db 376 GTCCTCATTAATAGAAATTAAGGCAATTTTGTGACGCGCGCGCGCGCTTTCCCA 435  
 QY 421 ATAGGAGGCGCAGTTTATCGGCGAGCTCTACTCTTCTTCTTATTTGGGTAAAGCCCTTTC 480  
 Db 436 ATAGGAGGCGCAGTTTATCGGCGAGCTCTACTCTTCTTCTTATTTGGGTAAAGCCCTTTC 495  
 QY 481 TGTTCCTGCGCAGTGTGCTGTGAGGCTGCGCGAGAACATAGTATAGATGATGATAC 540  
 Db 496 TGTTCCTGCGCAGTGTGCTGTGAGGCTGCGCGAGAACATAGTATAGATGATGATAC 555  
 QY 541 TTTTCATGAGAGATTAGCAAGGGAAGAAAACTATGCTGAGTGGAGTGTGTTTCA 600  
 Db 556 TTTTCATGAGAGATTAGCAAGGGAAGAAAACTATGCTGAGTGGAGTGTGTTTCA 615  
 QY 601 TCATATTAAGGAGAAATTTGCTCACTATGTGACAGTTTCTGGGACGTCTTAACTTT 660  
 Db 616 TCATATTAAGGAGAAATTTGCTCACTATGTGACAGTTTCTGGGACGTCTTAACTTT 675  
 QY 661 TATTGAGAGACTATCAATCATACAGATATTGTCAAAAAAAGACTAATAATA 720  
 Db 676 TATTGAGAGACTATCAATCATACAGATATTGTCAAAAAAAGACTAATAATA 725  
 QY 721 CAT 723  
 Db 736 CAT 738

RESULT 4  
 US-09-743-194-30  
 ; Sequence 30, Application US/09743194  
 ; Patent No. 6716601  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Belfield, Graham  
 ; APPLICANT: Oakley, Caroline  
 ; TITLE OF INVENTION: Compositions and Methods Utilizing Sequences for  
 ; FILE REFERENCE: 3526, 82543  
 ; CURRENT APPLICATION NUMBER: US/09/743,194  
 ; CURRENT FILING DATE: 2001-01-08  
 ; NUMBER OF SEQ ID NOS: 32  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 30  
 ; LENGTH: 850  
 ; TYPE: DNA  
 ; ORGANISM: Saccharomyces cerevisiae  
 US-09-743-194-30

Query Match 99.6%; Score 720.4; DB 4; Length 850;  
 Best Local Similarity 99.9%; Pred. No. 1.5e-228;  
 Matches 721; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTCATTAAGCAGCAACATCACTAGCTGCTCTAATAATACCTACGGAATA 60  
 Db 78 CTTTCATTAAGCAGCAACATCACTAGCTGCTCTAATAATACCTACGGAATA 137  
 QY 61 ACCATAAAGCAAGCAATCTACTTGGAGAAAGAGACGCTTGTAAAGGGGAT 120  
 Db 138 ACCATAAAGCAAGCAATCTACTTGGAGAAAGAGACGCTTGTAAAGGGGAT 197  
 QY 121 GGGGGCTAAGAGTCAATCTTTCTTTCTGCGGCTCGGAGCCCGGACCTCTCT 180  
 Db 198 GGGGGCTAAGAGTCAATCTTTCTTTCTGCGGCTCGGAGCCCGGACCTCTCT 257  
 QY 181 CTCGCCGACGATTTCTCTCTTCTATCTCTTTTATCTCATCCGCTGAAGCAAC 240

Db 258 CTCGCCGACGATTTCTCTCTTCTATCTCTTCTTATTTCCATCCCGGTGAAGCAAC 317  
 QY 241 GCATATGACTAATGCTGTGACATCTTCATGCTGTGACTTGTGTATCTCACAGT 300  
 Db 318 GCATATGACTAATGCTGTGACATCTTCATGCTGTGACTTGTGTATCTCACAGT 377  
 QY 301 GGTAACGGACCGCTGCTCGGAAACGGTTCCTTCTGTGACAAATCTAGAACAGGGGCTAC 360  
 Db 376 GGTAACGGACCGCTGCTCGGAAACGGTTCCTTCTGTGACAAATCTAGAACAGGGGCTAC 437  
 QY 361 GTCCTCATTAATAGAAATTAAGGCAATTTTGTGACGCGCGCGCGCGCTTTCCCA 420  
 Db 436 GTCCTCATTAATAGAAATTAAGGCAATTTTGTGACGCGCGCGCGCGCTTTCCCA 497  
 QY 421 ATAGGAGGCGCAGTTTATCGGCGAGCTCTACTCTTCTTCTTATTTGGGTAAAGCCCTTTC 480  
 Db 498 ATAGGAGGCGCAGTTTATCGGCGAGCTCTACTCTTCTTCTTATTTGGGTAAAGCCCTTTC 557  
 QY 481 TGTTCCTGCGCAGTGTGCTGTGAGGCTGCGCGAGAACATAGTATAGATGATGATAC 540  
 Db 558 TGTTCCTGCGCAGTGTGCTGTGAGGCTGCGCGAGAACATAGTATAGATGATGATAC 617  
 QY 541 TTTTCATGAGAGATTAGCAAGGGAAGAAAACTATGCTGAGTGGAGTGTGTTTCA 600  
 Db 618 TTTTCATGAGAGATTAGCAAGGGAAGAAAACTATGCTGAGTGGAGTGTGTTTCA 677  
 QY 601 TCATATTAAGGAGAAATTTGCTCACTATGTGACAGTTTCTGGGACGTCTTAACTTT 660  
 Db 678 TCATATTAAGGAGAAATTTGCTCACTATGTGACAGTTTCTGGGACGTCTTAACTTT 727  
 QY 661 TATTGAGAGACTATCAATCATACAGATATTGTCAAAAAAAGACTAATAATA 720  
 Db 738 TATTGAGAGACTATCAATCATACAGATATTGTCAAAAAAAGACTAATAATA 797  
 QY 721 CA 722  
 Db 798 AA 799

RESULT 5  
 US-08-232-463-14  
 ; Sequence 14, Application US/08232463  
 ; Patent No. 5670367  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DORNER, F.  
 ; APPLICANT: SCHEIFLINGER, F.  
 ; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
 ; NUMBER OF SEQUENCES: 52  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 1800 Diagonal Road, Suite 500  
 ; CITY: Alexandria  
 ; STATE: VA  
 ; COUNTRY: USA  
 ; ZIP: 22313-0299  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/232,463  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/935,313  
 ; FILING DATE:  
 ; APPLICATION NUMBER: EP 91 114 300.6  
 ; FILING DATE: 26-AUG-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BENT, Stephen A.  
 ; REGISTRATION NUMBER: 29,768



; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO: 16850  
 ; LENGTH: 16662  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-09-949-016-16850

Query Match  
 Best Local Similarity 54.1%; Score 36.6; DB 4; Length 16662;  
 Matches 75; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 89 GGAAGAGAAAGAGACAGCTTTGTAAGGGGATGGGGCTAAGATCATTCCTTTCTT 148  
 DB 4337 GGAAGAGAAAGAGACATGATATTAAGGGCTCAGACCTGCTTTTCTTCTTCTT 4338  
 QY 149 TCCCTTCGCGGTCGGAGCCGCGGACCTCTCTCCCGACAGATTCTTCTTCATAT 208  
 DB 4337 CTCCTC 4278  
 QY 209 CTCCTTTATTCCTATCC 227  
 DB 4277 CTCCTTTCTTTCTTTCTTCC 4259

RESULT 9  
 US-09-369-247-15  
 ; Sequence 15, Application US/09369247  
 ; Patent No. 6569992  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: 44 Human Secreted Proteins  
 ; FILE REFERENCE: P2024P1  
 ; CURRENT APPLICATION NUMBER: US/09/369,247  
 ; EARLIER FILING DATE: 1999-08-05  
 ; EARLIER APPLICATION NUMBER: 60/074,118  
 ; EARLIER FILING DATE: 1998-02-09  
 ; EARLIER APPLICATION NUMBER: 60/074,157  
 ; EARLIER FILING DATE: 1998-02-09  
 ; EARLIER APPLICATION NUMBER: 60/074,137  
 ; EARLIER FILING DATE: 1998-02-09  
 ; EARLIER APPLICATION NUMBER: 60/074,341  
 ; EARLIER FILING DATE: 1998-02-09  
 ; EARLIER APPLICATION NUMBER: 60/074,141  
 ; EARLIER FILING DATE: 1998-02-09  
 ; NUMBER OF SEQ ID NOS: 172  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO: 15  
 ; LENGTH: 815  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (406)  
 ; OTHER INFORMATION: n equals a,c,g, or c  
 ; NAME/KEY: SITE  
 ; LOCATION: (794)  
 ; OTHER INFORMATION: n equals a,c,g, or c  
 US-09-369-247-15

Query Match  
 Best Local Similarity 55.1%; Score 35.8; DB 4; Length 815;  
 Matches 70; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 594 TTTTCATCATATTAAGGAGAAATTTGCTCACTATGTGACAGTTTCTGGAGCTCT 653  
 DB 665 TATCTGTATTTAAATGTAACATTTATTTGATCAGTCACTGATTTAGTTAAATAGT 724  
 QY 654 TAACTTTATGAGAGACTATCAATATCATACAGATTTGTCAAAAAAAGAGCTA 713

DB 725 CTTCGCTATGACAGAGTGTAATAATTAATTAATTTTAAATTAATAAAAAA 784  
 QY 714 ATATATA 720  
 DB 785 AAAAAA 791

RESULT 10  
 US-09-949-016-16144  
 ; Sequence 16144, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; FILE REFERENCE: CL001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO: 16144  
 ; LENGTH: 191433  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-09-949-016-16144

Query Match  
 Best Local Similarity 59.4%; Score 35.4; DB 4; Length 191433;  
 Matches 60; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 104 AGCCTGTAAAGGGGATGGGGCTAAGATCATTCCTTTCTTCCCTGCGGCTCG 163  
 DB 41249 ACAGTGTTAAGAGAGAGAGTAGAATCAGTTGTTCAAGTTCTTTCTTCCAGGTAA 41308  
 QY 164 GACCGGAGACCCCTCTCCCGCAAGATTCTTCTTTC 204  
 DB 41309 TCCCGCCACCCCGCGCTGCTCCCTTCCGCAATTTC 41349

RESULT 11  
 US-09-949-016-15419/C  
 ; Sequence 15419, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CL001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO: 15419  
 ; LENGTH: 212449  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)-(212449)  
 ; OTHER INFORMATION: n = A,T,C or G

QY 594 TTTTCATCATATTAAGGAGAAATTTGCTCACTATGTGACAGTTTCTGGAGCTCT 653  
 DB 665 TATCTGTATTTAAATGTAACATTTATTTGATCAGTCACTGATTTAGTTAAATAGT 724  
 QY 654 TAACTTTATGAGAGACTATCAATATCATACAGATTTGTCAAAAAAAGAGCTA 713

US-09-949-016-15419

Query Match 4.9%; Score 35.4; DB 4; Length 212449;  
Best Local Similarity 57.8%; Pred. No. 13;  
Matches 63; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 123 GGGCTAAGAGTATCATCTTCTTCCCTTGGCGGTCCGGACCCCGGACCCCTCTCT 182  
DB 124084 GGTCTCATATTCATTCATTCCTCTTCCCTTCCCTCTCTCTCTCTCTCTCTCC 124025  
QY 183 CCCCCGACATTTCTTCTTCAATCTCTTATCTCTATCCGCTT 231  
DB 124024 CTCCTTCTCTTCTTCTTCTCTCTCTTCTCTTCTCTCTCTCTCTCTCTCT 123976

RESULT 12

US-09-949-016-17592/c  
Sequence 17592, Application US/09949016  
Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

FILE REFERENCE: CL001307

CURRENT APPLICATION NUMBER: US/09/949,016

PRIOR FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

PRIOR FILING DATE: 2000-09-08

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 17592

LENGTH: 72992

TYPE: DNA

ORGANISM: Human

US-09-949-016-17592

Query Match 4.9%; Score 35.2; DB 4; Length 72992;  
Best Local Similarity 60.4%; Pred. No. 7.9;  
Matches 58; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 132 AGTATTCACCTTTTCTTCCCTTGGCGGTCCGGACCCCGGACCCCTCTCTCCCGACG 191  
DB 72622 AGCTCTTCTTCTTCTTCTTCTTCCCTTCCCTCCCTCCCTCCCTCCCTCCCTCC 72563  
QY 192 ATTCTTCTTCTTCAATCTCTTCTTATCTCTATCC 227  
DB 72562 CTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCC 72527

RESULT 13

US-09-949-016-14133/c  
Sequence 14133, Application US/09949016  
Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

FILE REFERENCE: CL001307

CURRENT APPLICATION NUMBER: US/09/949,016

PRIOR FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

PRIOR FILING DATE: 2000-09-08

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 14133

LENGTH: 98962

TYPE: DNA

ORGANISM: Human

FEATURE:

NAME/KEY: misc.feature

LOCATION: (1)...(98962)

OTHER INFORMATION: n = A,T,C or G

US-09-949-016-14133

Query Match 4.8%; Score 34.8; DB 4; Length 98962;  
Best Local Similarity 53.7%; Pred. No. 13;  
Matches 72; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 84 TACTTGGAGGAAAGAGACAGCTTGTAAAGGGGATGGGGGCTAAGATCATTCACCT 143  
DB 78525 TCTGGGAGGAGCAGTGGGTGGTGTAGTGGGAGGAGGCTTACACCCCAACATTT 78466  
QY 144 TCTTTTCCCTTGGCGGTCCGACCCCGGACCCCTCTCTCCCGACGATTTCTCTTT 203  
DB 78465 TCCCAACCTTGGCCACACACCCCTGGAAACCAGCTCTAGCACTATCTTTT 78406  
QY 204 CATATCTTCTTTT 217  
DB 78405 TTTTCTTTTCTTTT 78392

RESULT 14

US-09-949-016-17100/c  
Sequence 17100, Application US/09949016  
Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

FILE REFERENCE: CL001307

CURRENT APPLICATION NUMBER: US/09/949,016

PRIOR FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

PRIOR FILING DATE: 2000-09-08

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 17100

LENGTH: 102884

TYPE: DNA

ORGANISM: Human

FEATURE:

NAME/KEY: misc.feature

LOCATION: (1)...(102884)

OTHER INFORMATION: n = A,T,C or G

US-09-949-016-17100

Query Match 4.8%; Score 34.8; DB 4; Length 102884;  
Best Local Similarity 53.7%; Pred. No. 13;  
Matches 72; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 84 TACTTGGAGGAAAGAGACAGCTTGTAAAGGGGATGGGGGCTAAGATCATTCACCT 143  
DB 78525 TCTGGGAGGAGCAGTGGGTGGTGTAGTGGGAGGAGGCTTACACCCCAACATTT 78466  
QY 144 TCTTTTCCCTTGGCGGTCCGACCCCGGACCCCTCTCTCCCGACGATTTCTCTTT 203  
DB 78465 TCCCAACCTTGGCCACACACCCCTGGAAACCAGCTCTAGCACTATCTTTT 78406  
QY 204 CATATCTTCTTTT 217  
DB 78405 TTTTCTTTTCTTTT 78392

RESULT 15



```

US-09-949-016-26160/C
; Sequence 26160, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26160
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-26160

```

```

Query Match      4.8%; Score 34.4; DB 4; Length 601;
Best Local Similarity 54.8%; Pred. No. 0.82;
Matches 68; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```

```

QY 25 CACATAGACTGCGTCATTAATAATACCTACGAAAAACATAAGAGCAAGGATACCT 84
   |||||
DB 409 CAAAGAGAGAGACTCTCAAAAACAAAACAAAACAAAACAAAACAAAACGAG 350
   |||||
QY 85 ACTTGAAGGAAAAAGAGCAGCGCTGTGAAGGGGATGGGGCTAAGAGTCATTCACTT 144
   |||||
DB 349 ACTGGCCAGAAACAGTAATAGAGGCAAAAAGAGGTGTGATGAAGAAYCTATATATGT 290
   |||||
QY 145 CTTT 148
   |||
DB 289 CATT 286
   |||

```

Search completed: February 7, 2005, 23:20:25  
 Job time : 180 secs

**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2005, 22:23:48 ; Search time 526 Seconds  
(without alignments)  
7911.410 Million cell updates/sec

Title: US-10-776-213-2  
Perfect score: 1 cttcgattagcagcagcagc.....aaaagactaataacat 723  
Sequence: 723

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 4313806 seqs, 2877871033 residues  
Total number of hits satisfying chosen parameters: 8627612

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 19: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 20: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 21: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 22: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	723	100.0	723	US-10-776-213-2	Sequence 2, Appl1
2	723	100.0	11427	US-10-776-213-24	Sequence 24, Appl1
3	723	100.0	13073	US-10-776-213-20	Sequence 20, Appl1
4	720.4	99.6	850	US-10-776-213-30	Sequence 30, Appl1
5	497.4	68.8	680	US-10-451-467A-317	Sequence 317, App
6	411	5.7	1024	US-10-123-155-198	Sequence 198, App
7	411	5.7	1024	US-10-146-731-198	Sequence 198, App
8	411	5.7	1024	US-10-140-472-198	Sequence 198, App
9	411	5.7	1024	US-10-141-761-198	Sequence 198, App
10	411	5.7	1024	US-10-142-885-198	Sequence 198, App
11	411	5.7	1024	US-10-158-790-198	Sequence 198, App

C 12	41	5.7	1024	17	US-10-137-871-198	Sequence 198, App
C 13	41	5.7	1024	17	US-10-140-923-198	Sequence 198, App
C 14	41	5.7	1024	17	US-10-141-756-198	Sequence 198, App
C 15	41	5.7	1024	17	US-10-141-759-198	Sequence 198, App
C 16	41	5.7	1024	17	US-10-140-805-198	Sequence 198, App
C 17	41	5.7	1024	17	US-10-140-864-198	Sequence 198, App
C 18	41	5.7	1024	17	US-10-142-426-198	Sequence 198, App
C 19	41	5.7	2561	17	US-10-197-824-20	Sequence 20, Appl1
C 20	39.4	5.4	409	17	US-10-424-599-136845	Sequence 136845, App
C 21	38.6	5.3	2311	18	US-10-466-531-23	Sequence 23, Appl1
C 22	38.4	5.3	544	14	US-10-066-543-1775	Sequence 1775, App
C 23	38.4	5.3	2355	17	US-10-424-599-29921	Sequence 29921, A
C 24	38.2	5.3	1635	17	US-10-424-599-129944	Sequence 129944, App
C 25	37.8	5.2	7736	9	US-09-764-846-333	Sequence 333, App
C 26	37.8	5.2	7736	14	US-10-091-483-333	Sequence 333, App
C 27	37.4	5.2	386	18	US-10-357-930-4385	Sequence 4385, App
C 28	37.2	5.1	436	17	US-10-424-599-96559	Sequence 96559, A
C 29	37.2	5.1	685	13	US-10-027-632-106894	Sequence 106894, App
C 30	37.2	5.1	685	13	US-10-027-632-106895	Sequence 106895, App
C 31	37.2	5.1	685	17	US-10-027-632-106894	Sequence 106894, App
C 32	37.2	5.1	685	17	US-10-027-632-106895	Sequence 106895, App
C 33	37.2	5.1	704	13	US-10-027-632-25182	Sequence 25182, A
C 34	37.2	5.1	704	13	US-10-027-632-25183	Sequence 25183, A
C 35	37.2	5.1	704	17	US-10-027-632-25182	Sequence 25182, A
C 36	37.2	5.1	121	17	US-10-027-632-25183	Sequence 25183, A
C 37	37.2	5.1	1121	17	US-10-424-599-94399	Sequence 94399, A
C 38	37	5.1	366	9	US-09-764-877-819	Sequence 819, App
C 39	37	5.1	366	17	US-10-242-515-819	Sequence 819, App
C 40	37	5.1	390	9	US-09-960-352-1976	Sequence 1976, App
C 41	37	5.1	400	11	US-09-969-034-2883	Sequence 2883, App
C 42	37	5.1	484	18	US-10-425-115-4593	Sequence 4593, App
C 43	36.8	5.1	700	17	US-10-027-632-168936	Sequence 168936, App
C 44	36.8	5.1	700	17	US-10-027-632-168936	Sequence 168936, App
C 45	36.6	5.1	1980090	18	US-10-719-993-6815	Sequence 6815, App

# ALIGNMENTS

RESULT 1  
US-10-776-213-2  
; Sequence 2, Application US/10776213  
; Publication NO. US20040142478A1  
; GENERAL INFORMATION:  
; APPLICANT: Astrazenca AB  
; TITLE OF INVENTION: Compositions and Methods Utilizing Sequences for Controlling Nucleic  
; TITLE OF INVENTION: Acid Expression  
; FILE REFERENCE: 3526.82543  
; CURRENT APPLICATION NUMBER: US/10/776, 213  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 723  
; TYPE: DNA  
; ORGANISM: Saccharomyces cerevisiae  
; US-10-776-213-2

Query Match 100.0%; Score 723; DB 18; Length 723;  
Best Local Similarity 100.0%; Pred. No. 5.5e-217; Indels 0; Gaps 0;  
Matches 723; Conservative 0; Mismatches 0;

QY	1	CTTCGATTAGCAGCAGACATCATGACTGGCTATTAATAATACATACGGAANA	60
DB	1	CTTCGATTAGCAGCAGACATCATGACTGGCTATTAATAATACATACGGAANA	60
QY	61	ACCATTAAGAGAAACGATCTACTTGAAGAAAGAGAGACGCTTGAAGGGGAT	120
DB	61	ACCATTAAGAGAAACGATCTACTTGAAGAAAGAGAGACGCTTGAAGGGGAT	120
QY	121	GGGGGCTAAGAGATCTTCTTCCCTCGCGGTCGGACCCCGGACCCCTCTCT	180
DB	121	GGGGGCTAAGAGATCTTCTTCCCTCGCGGTCGGACCCCGGACCCCTCTCT	180

```
QY 181 CTCCCGGACGATTTCTTCTTTGATATCTTCTTTATTCCTTAATCCCGTTGAAGCAACC 240
Db 181 CTCCCGGACGATTTCTTCTTTGATATCTTCTTTATTCCTTAATCCCGTTGAAGCAACC 240
QY 241 GCACATATGACTAATAATGGTGTGACATCTCCATGGCTGTGACTGTGTGATCTCAAGT 300
Db 241 GCACATATGACTAATAATGGTGTGACATCTCCATGGCTGTGACTGTGTGATCTCAAGT 300
QY 301 GGTAAAGGACCGTGGCTCGGAAAACGTTCTTCTGTGACAAATTTAGAACAGGGGCTTACA 360
Db 301 GGTAAAGGACCGTGGCTCGGAAAACGTTCTTCTGTGACAAATTTAGAACAGGGGCTTACA 360
QY 361 GTCGTAAATATAGAAATATAAGCGCATTTTGTCTAGAGCGCGCGCGCGCGCTTTTCCA 420
Db 361 GTCGTAAATATAGAAATATAAGCGCATTTTGTCTAGAGCGCGCGCGCGCGCTTTTCCA 420
QY 421 ATAGGAGGCGCAGTTTATCGCGGAGCTCTACTTCTTCTTAATTTGGGTAAAGCCCTTTTC 480
Db 421 ATAGGAGGCGCAGTTTATCGCGGAGCTCTACTTCTTCTTAATTTGGGTAAAGCCCTTTTC 480
QY 481 TGTTTTGGCCAGTGTGTCTGTGACAGCTCGCGCGGAGAACATAGTAAAGGATGTAAAC 540
Db 481 TGTTTTGGCCAGTGTGTGTCTGTGACAGCTCGCGCGGAGAACATAGTAAAGGATGTAAAC 540
QY 541 TTTGATGAGAGAAATTAAGCAAGCGGAAAAAACTATGGCTAGCTGGAGTGTTTTCAA 600
Db 541 TTTGATGAGAGAAATTAAGCAAGCGGAAAAAACTATGGCTAGCTGGAGTGTTTTCAA 600
QY 601 TCATATATTAAGGAGAAATTTGTCTCACTATGTGACAGTTTCTGGAGCTCTTAACCTT 660
Db 601 TCATATATTAAGGAGAAATTTGTCTCACTATGTGACAGTTTCTGGAGCTCTTAACCTT 660
QY 661 TATTGCAAGAGACTATCAATATCATACATATATGTCAAAAAAAAGACTAATATTA 720
Db 661 TATTGCAAGAGACTATCAATATCATACATATATGTCAAAAAAAAGACTAATATTA 720
QY 721 CAT 723
Db 721 CAT 723

RESULT 2
US-10-776-213-24
; Sequence 24, Application US/10776213
; Publication No. US20040142478A1
; GENERAL INFORMATION:
; APPLICANT: Astrazeneca AB
; TITLE OF INVENTION: Compositions and Methods Utilizing Sequences for Controlling Nucleic Acid Expression
; FILE REFERENCE: 3526.82543
; CURRENT APPLICATION NUMBER: US/10/776,213
; CURRENT FILING DATE: 2004-02-12
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 11427
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-10-776-213-24
```

Query Match 100.0%; Score 723; DB 18; Length 11427;  
Best Local Similarity 100.0%; Pred. No. 2,5e-216;  
Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CTTTCGATTAGACGACACATCACTATGACTGCGTCAATAAATATACACTAGGAAAA 60
Db 15 CTTTCGATTAGACGACACATCACTATGACTGCGTCAATAAATATACACTAGGAAAA 74
QY 61 ACCATAAAGAGCAAGGATACCTACTTGGAGAGAAAAAGAGACGCTGTAAAGGGGAT 120
Db 75 ACCATAAAGAGCAAGGATACCTACTTGGAGAGAAAAAGAGACGCTGTAAAGGGGAT 134
```

```
QY 121 GGGGGCTAAGAGTCAATCACTTTCTTTTCCCTTCGCGGTCCGAGCCCGGACCCCTCT 180
Db 135 GGGGGCTAAGAGTCAATCACTTTCTTTTCCCTTCGCGGTCCGAGCCCGGACCCCTCT 194
QY 181 CTCCCGGACGATTTCTTCTTTGATATCTTCTTTATTCCTTAATCCCGTTGAAGCAACC 240
Db 195 CTCCCGGACGATTTCTTCTTTGATATCTTCTTTATTCCTTAATCCCGTTGAAGCAACC 254
QY 241 GCACATATGACTAATAATGGTGTGACATCTCCATGGCTGTGACTGTGTGATCTCAAGT 300
Db 255 GCACATATGACTAATAATGGTGTGACATCTCCATGGCTGTGACTGTGTGATCTCAAGT 314
QY 301 GGTAAAGGACCGTGGCTCGGAAAACGTTCTTCTGTGACAAATTTAGAACAGGGGCTTACA 360
Db 315 GGTAAAGGACCGTGGCTCGGAAAACGTTCTTCTGTGACAAATTTAGAACAGGGGCTTACA 374
QY 361 GTCGTAAATATAGAAATATAAGCGCATTTTGTCTAGAGCGCGCGCGCGCGCTTTTCCA 420
Db 375 GTCGTAAATATAGAAATATAAGCGCATTTTGTCTAGAGCGCGCGCGCGCGCTTTTCCA 434
QY 421 ATAGGAGGCGCAGTTTATCGCGGAGCTCTACTTCTTCTTAATTTGGGTAAAGCCCTTTTC 480
Db 435 ATAGGAGGCGCAGTTTATCGCGGAGCTCTACTTCTTCTTAATTTGGGTAAAGCCCTTTTC 494
QY 481 TGTTTTGGCCAGTGTGTCTGTGACAGCTCGCGCGGAGAACATAGTAAAGGATGTAAAC 540
Db 495 TGTTTTGGCCAGTGTGTGTCTGTGACAGCTCGCGCGGAGAACATAGTAAAGGATGTAAAC 554
QY 541 TTTGATGAGAGAAATTAAGCAAGCGGAAAAAACTATGGCTAGCTGGAGTGTTTTCAA 600
Db 555 TTTGATGAGAGAAATTAAGCAAGCGGAAAAAACTATGGCTAGCTGGAGTGTTTTCAA 614
QY 601 TCATATATTAAGGAGAAATTTGTCTCACTATGTGACAGTTTCTGGAGCTCTTAACCTT 660
Db 615 TCATATATTAAGGAGAAATTTGTCTCACTATGTGACAGTTTCTGGAGCTCTTAACCTT 674
QY 661 TATTGCAAGAGACTATCAATATCATACATATATGTCAAAAAAAAGACTAATATTA 720
Db 675 TATTGCAAGAGACTATCAATATCATACATATATGTCAAAAAAAAGACTAATATTA 734
QY 721 CAT 723
Db 735 CAT 737
```

RESULT 3  
US-10-776-213-20  
; Sequence 20, Application US/10776213  
; Publication No. US20040142478A1  
; GENERAL INFORMATION:  
; APPLICANT: Astrazeneca AB  
; TITLE OF INVENTION: Compositions and Methods Utilizing Sequences for Controlling Nucleic Acid Expression  
; FILE REFERENCE: 3526.82543  
; CURRENT APPLICATION NUMBER: US/10/776,213  
; CURRENT FILING DATE: 2004-02-12  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 13073  
; TYPE: DNA  
; ORGANISM: Saccharomyces cerevisiae  
US-10-776-213-20

Query Match 100.0%; Score 723; DB 18; Length 13073;  
Best Local Similarity 100.0%; Pred. No. 2,7e-216;  
Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CTTTCGATTAGACGACACATCACTATGACTGCGTCAATAAATATCACTAGGAAAA 60
Db 16 CTTTCGATTAGACGACACATCACTATGACTGCGTCAATAAATATCACTAGGAAAA 75
QY 61 ACCATAAAGAGCAAGGATACCTACTTGGAGAGAAAAAGAGACGCTGTAAAGGGGAT 120
```







Db 534 ..TTCMWRT.WH.BYH.HAT..B..SB.DSST.CH..KB.D.BHA..Y.M.KMWC.R.RH 475  
Qy 527 ATAAAGGATGTAACCTTCATGAGAGAAATAGCAAGCGGAAAAAACTATGCTAGCTGG 586  
Db 474 ACB...RDT.C.M.WBH.WWB.AB.HC.W.DG..DB.BKABH..RS.SSBS.H...Y..B 415  
Qy 587 GAGTGTTCATCATATATAAAGGAGAAATGCTGCTACATATGACAGTTCTGG 646  
Db 414 ..S.CWN.KTTR.DW.BT..H.M..SR.BM.RH.R...YMA.D.TRHSCYV.TH.YTR 355  
Qy 647 GAGCTTTAATCTTTATTTGACAGAGACTATCAATCATACAGATATTTGTCAAAAAAA 706  
Db 354 .YBBBCSMAA.YRA.NG.TWT.RRASK..A.S.DHABWCH...BSA.B.KNS.S.RAT. 295  
Qy 707 AAGACTAATAATAC 721  
Db 294 DW.S.M.ATMR...H 280

RESULT 9  
US-10-141-761-198/c

/ Sequence 198, Application US/10141761  
/ Publication No. US20030148432A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Baker, Kevin P.  
/ APPLICANT: Beresini, Maureen  
/ APPLICANT: Deforge, Laura  
/ APPLICANT: Desnoyers, Luc  
/ APPLICANT: Filvaroff, Ellen  
/ APPLICANT: Gao, Wei-Qiang  
/ APPLICANT: Gerritsen, Mary E.  
/ APPLICANT: Goddard, Audrey  
/ APPLICANT: Godowski, Paul J.  
/ APPLICANT: Gurney, Austin L.  
/ APPLICANT: Sherwood, Steven  
/ APPLICANT: Smith, Victoria  
/ APPLICANT: Stewart, Timothy A.  
/ APPLICANT: Tumas, Daniel  
/ APPLICANT: Watanabe, Colin K  
/ APPLICANT: Wood, William  
/ APPLICANT: Zhang, Zemin  
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
/ FILE REFERENCE: P330R1C198  
/ CURRENT APPLICATION NUMBER: US/10/141,761  
/ CURRENT FILING DATE: 2002-05-08  
/ Prior Application removed - See Palm or File Wrapper  
/ NUMBER OF SEQ ID NOS: 550  
/ SEQ ID NO 198  
/ LENGTH: 1024  
/ TYPE: PRT  
/ ORGANISM: Homo Sapien  
US-10-141-761-198

Query Match 5.7%; Score 41; DB 15; Length 1024;  
Best Local Similarity 7.7%; Pred. No. 0.088;  
Matches 52; Conservative 194; Mismatches 429; Indels 0; Gaps 0;

Qy 47 TACACTACGGAAGAAACATTAAGAGCAAGCAACTACTATGGAAGAAAGAGACACG 106  
Db 954 T.CRYAN..MMWIV..MWT.HYCD.NYTB.D.BH.TH.H.BB..SN.S.N..YN..NYHS 895  
Qy 107 CTTGTAAGGGGAGTGGGGCTAAGAATCATCTTTCTTTCCCTGCGGCGCGAGC 166  
Db 894 GA.T...WM.SBTAASD.MKW...WBH.H..SBGAGR.BYBN...YVYNS.WHS..A 835  
Qy 167 CCGGAGCCCTCTCTCCCGGACGATTTCTTCTTCATATCTTCTTTATCTTCTATC 226  
Db 834 ST..DSDMNC.SGT.K.B.YY...D.RST.B.SCN.NM.WC.M.RAAB.B.M..R.N..N 775  
Qy 227 CCGTTGAAGCAACGACATGACTAAATGCTGCTGACATCTCCAGATGGGTGACTTGT 286  
Db 774 RN.KYH..MH..R.TYSTIDW..HM..S.RY.....N...RCTYT.S.THH..CTYNS 715

Qy 287 GTGTATCTCAGATGGTAACGCGACCGGTGCTGGAAAGCTTCCTTCGATATTCYA 346  
Db 714 BS.B.DNY...H..YMWY.HRNBY.RCA.N.NC..MSCWH.RA.YDD.SMNSBW..T.S 655  
Qy 347 GAACAGGGGCTACAGTCTCGATATATAGATATAGCGCAATTTTCTAGCGCGCGG 406  
Db 654 SBDSYNCB.A...W.RSNM..M.TMAMTS..HR..D..A...YN.TAANC..A.B.RCK 595  
Qy 407 GCGCCGTTTCCCATATAGGAGGCGAGCTTATCGGCGGAGCTCTCTCTTTTG 466  
Db 594 ..AM.HKB..NBRMCNM.T.S.AMW.HNYTNCMD.A.RH..CY..BSDCK.NT.KY 535  
Qy 467 GGTAAAGCCCTTCTCTTTTTCGCGAGGTGCTGCGAGCTGCGCGGAGCAATAGTG 526  
Db 534 ..TTCMWRT.WH.BYH.HAT..B..SB.DSST.CH..KB.D.BHA..Y.M.KMWC.R.RH 475  
Qy 527 ATAAAGGATGTAACCTTCATGAGAGAAATAGCAAGCGGAAAAAACTATGCTAGCTGG 586  
Db 474 ACB...RDT.C.M.WBH.WWB.AB.HC.W.DG..DB.BKABH..RS.SSBS.H...Y..B 415  
Qy 587 GAGTGTTCATCATATATAAAGGAGAAATGCTGCTACATATGACAGTTCTGG 646  
Db 414 ..S.CWN.KTTR.DW.BT..H.M..SR.BM.RH.R...YMA.D.TRHSCYV.TH.YTR 355  
Qy 647 GAGCTTTAATCTTTATTTGACAGAGACTATCAATCATACAGATATTTGTCAAAAAAA 706  
Db 354 .YBBBCSMAA.YRA.NG.TWT.RRASK..A.S.DHABWCH...BSA.B.KNS.S.RAT. 295  
Qy 707 AAGACTAATAATAC 721  
Db 294 DW.S.M.ATMR...H 280

RESULT 10  
US-10-142-885-198/c

/ Sequence 198, Application US/10142885  
/ Publication No. US20030157604A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Baker, Kevin P.  
/ APPLICANT: Beresini, Maureen  
/ APPLICANT: Deforge, Laura  
/ APPLICANT: Desnoyers, Luc  
/ APPLICANT: Filvaroff, Ellen  
/ APPLICANT: Gao, Wei-Qiang  
/ APPLICANT: Gerritsen, Mary E.  
/ APPLICANT: Goddard, Audrey  
/ APPLICANT: Godowski, Paul J.  
/ APPLICANT: Gurney, Austin L.  
/ APPLICANT: Sherwood, Steven  
/ APPLICANT: Smith, Victoria  
/ APPLICANT: Stewart, Timothy A.  
/ APPLICANT: Tumas, Daniel  
/ APPLICANT: Watanabe, Colin K  
/ APPLICANT: Wood, William  
/ APPLICANT: Zhang, Zemin  
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
/ FILE REFERENCE: P330R1C248  
/ CURRENT APPLICATION NUMBER: US/10/142,885  
/ CURRENT FILING DATE: 2002-05-10  
/ Prior Application removed - See File Wrapper or Palm  
/ NUMBER OF SEQ ID NOS: 550  
/ SEQ ID NO 198  
/ LENGTH: 1024  
/ TYPE: PRT  
/ ORGANISM: Homo Sapien  
US-10-142-885-198

Query Match 5.7%; Score 41; DB 16; Length 1024;  
Best Local Similarity 7.7%; Pred. No. 0.088;  
Matches 52; Conservative 194; Mismatches 429; Indels 0; Gaps 0;



```

QY 47 TACACTACGAAAACCATTAAGAGCAAGCACTACTACTTGGAGAAAGAGACACG 106
DB 954 T..CRYAN..MMWHY..MMT.HYCD..NYTBD..BH..TH..H..BB..SN..S..N..YN..NYHHS 895
QY 107 CTGTGAAGGGGCTAGGAGGCTAAGAGTCACTTCTTCCCTCCGCGCGAC 166
DB 894 GA..T...MM..SBTAASD..MKW..WBH..H..SBGAGR..HYEN...YNNSS..MHS..A 835
QY 167 CCGGAGACCCCTCTCTCCCGGACGATTTCTTCTTCAATATCTTCTTTATCTTATC 226
DB 834 ST..DEDMNC..SGT..K..B..Y...D..RST..B..SCN..NNM..WC..M..RAATB..M..R..N..N 775
QY 227 CCCTTGAAGAGCCGACCTATGACTAATAGTGCTGACATCTCCATGCTGACTGT 286
DB 774 RN..KYN..MH..R..TYSTTDM...HM..S..RY.....N...RCTY..S..THH..CTYNS 715
QY 287 GTGTATCTCAGATGTGTAACGCGACCGTGGCTCGGAAACGTTCTTCTTGAACAATCTA 346
DB 714 BS..B..DNY...H...YWNV..HRMBY..RCA..N..NC..WSCNH..RA..YVD..SMNSBW..T..S 655
QY 347 GAACAGGGGCTACAGTCTCGATATAGATTAATAGAGCAATTTTGTAGCGCGCGCG 406
DB 654 SBDSYNCB..A...W..RSNN..M..TWAITS..HR..D..A...YN..TAANC..A..B..RCK 595
QY 407 GCGCCGCTTCCCAATAGGAGGCGGCTTATCGCGGAGCTTACTTCTTCTTATTTG 466
DB 594 ..AM..HKB...NBRMCNVM..T..S..ANW..HNYTTCRMD..A..RH..CY..BSDCK..NT..KY 535
QY 467 GGTAAAGCCCTTCTGTCTTCTGCGGAGTGTGCTGCGAGCGCGGAGAACATAGTG 526
DB 534 ..TTCCMRT..WH..BYH..HAT..B..SB..DSST..CH..KB..D..BHA..Y..M..KXWC..R..RH 475
QY 527 ATAAAGGATGTAACTTTCGATGAGAGATTAAGAGGCAAGGAAACCTATGCTAGCTGG 586
DB 474 ACB...RDT..C..M..WBH..WBH..AB..HC..W..DG...DB..BKABH...RS..SSBS..H...Y..B 415
QY 587 GAGTTGTTTTCATCATATAAAGGAGAAATTTGCTCAGTATGTGACATTTCTGG 646
DB 414 ..S..CMN..KTRR..DM..BT..H..M..SR..BM..RH..R...YMA..D..TRHSCYV..TH..YTR 355
QY 647 GACGCTTAACTTTTATGCGAGAGACTATCAATCAATACAGATATGTCAAAAAAAA 706
DB 354 ..YBBSGMAAA..YRA..NG..TWT..RRASK..A..S..DHABWCH...BSA..B..KNS..S..RAT 295
QY 707 AAGACTAATATAAC 721
DB 294 DM..S..M..ATWR...H 280

```

RESULT 11  
US-10-158-790-198/c  
Sequence 198, Application US/10158790  
Publication No. US20030180879A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Geriltsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
TITLE OF INVENTION: ACIDS ENCODING THE SAME

```

FILE REFERENCE: P3330R1C448
CURRENT APPLICATION NUMBER: US/10/158, 790
CURRENT FILING DATE: 2002-05-30
Prior Application removed - See file wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 198
LENGTH: 1024
TYPE: PRT
ORGANISM: Homo Sapien
US-10-158-790-198

Query Match
5.7%; Score 41; DB 16; Length 1024;
Best Local Similarity 7.7%; Pred. No. 0.008;
Matches 52; Conservative 194; Mismatches 429; Indels 0; Gaps 0;

QY 47 TACACTACGAAAACCATTAAGAGCAAGCACTACTACTTGGAGAAAGAGACACG 106
DB 954 T..CRYAN..MMWHY..MMT.HYCD..NYTBD..BH..TH..H..BB..SN..S..N..YN..NYHHS 895
QY 107 CTGTGAAGGGGCTAGGAGGCTAAGAGTCACTTCTTCCCTCCGCGCGAC 166
DB 894 GA..T...MM..SBTAASD..MKW..WBH..H..SBGAGR..HYEN...YNNSS..MHS..A 835
QY 167 CCGGAGACCCCTCTCTCCCGGACGATTTCTTCTTCAATATCTTCTTTATCTTATC 226
DB 834 ST..DEDMNC..SGT..K..B..Y...D..RST..B..SCN..NNM..WC..M..RAATB..M..R..N..N 775
QY 227 CCCTTGAAGAGCCGACCTATGACTAATAGTGCTGACATCTCCATGCTGACTGT 286
DB 774 RN..KYN..MH..R..TYSTTDM...HM..S..RY.....N...RCTY..S..THH..CTYNS 715
QY 287 GTGTATCTCAGATGTGTAACGCGACCGTGGCTCGGAAACGTTCTTCTTGAACAATCTA 346
DB 714 BS..B..DNY...H...YWNV..HRMBY..RCA..N..NC..WSCNH..RA..YVD..SMNSBW..T..S 655
QY 347 GAACAGGGGCTACAGTCTCGATATAGATTAATAGAGCAATTTTGTAGCGCGCGCG 406
DB 654 SBDSYNCB..A...W..RSNN..M..TWAITS..HR..D..A...YN..TAANC..A..B..RCK 595
QY 407 GCGCCGCTTCCCAATAGGAGGCGGCTTATCGCGGAGCTTACTTCTTCTTATTTG 466
DB 594 ..AM..HKB...NBRMCNVM..T..S..ANW..HNYTTCRMD..A..RH..CY..BSDCK..NT..KY 535
QY 467 GGTAAAGCCCTTCTGTCTTCTGCGGAGTGTGCTGCGAGCGCGGAGAACATAGTG 526
DB 534 ..TTCCMRT..WH..BYH..HAT..B..SB..DSST..CH..KB..D..BHA..Y..M..KXWC..R..RH 475
QY 527 ATAAAGGATGTAACTTTCGATGAGAGATTAAGAGGCAAGGAAACCTATGCTAGCTGG 586
DB 474 ACB...RDT..C..M..WBH..WBH..AB..HC..W..DG...DB..BKABH...RS..SSBS..H...Y..B 415
QY 587 GAGTTGTTTTCATCATATAAAGGAGAAATTTGCTCAGTATGTGACATTTCTGG 646
DB 414 ..S..CMN..KTRR..DM..BT..H..M..SR..BM..RH..R...YMA..D..TRHSCYV..TH..YTR 355
QY 647 GACGCTTAACTTTTATGCGAGAGACTATCAATCAATACAGATATGTCAAAAAAAA 706
DB 354 ..YBBSGMAAA..YRA..NG..TWT..RRASK..A..S..DHABWCH...BSA..B..KNS..S..RAT 295
QY 707 AAGACTAATATAAC 721
DB 294 DM..S..M..ATWR...H 280

```

RESULT 12  
US-10-137-871-198/c  
Sequence 198, Application US/10137871  
Publication No. US20030207350A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Desnoyers, Luc

```
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerltsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C153
CURRENT APPLICATION NUMBER: US/10/137,871
CURRENT FILING DATE: 2002-05-03
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 198
LENGTH: 1024
TYPE: PRT
ORGANISM: Homo Sapien
US-10-137-871-198
```

Query Match 5.7%; Score 41; DB 17; Length 1024;

Best Local Similarity 7.7%; Pred. No. 0.086; Matches 52; Conservative 194; Mismatches 429; Indels 0; Gaps 0;

```
47 TACACTACGAAAAACCATTAAGAGCAAGATCTACTTGGAGAAAGAGAGCAG 106
954 T.CRYAN..MMMH..MWT.HYCD.NYTD..BH.TH.H.BB.SN.S.N..YN..NYHS 895
107 CTGTGAAGGGAGTGGGGCTAAGAGTCACTTCTTTCCCTCGCGCGGAC 166
894 GA.T...MM.SBTAASD.MKW..WBH.H..SBQAGR.HYBN...YNYNS.WHS..A 835
167 CCGGAGCCCTCTCTCCCGCAGATTTCTCTTCAATATCTCTTATTCATC 226
834 ST..DGMNC.SGT.K.B.YY..D.RST.B.SCN.NNM.WC.M.RATB.M..R.N..N 775
227 CCGTTGAAGCAACCGCATATGACTAATGCTGACATCTCCATGCTGACTGT 286
774 RN.KYH..MH..R.TYSTTDW...HM..S.RY.....N..RCTYT.S.THH..CTYNS 715
287 GTGTATCTACAGTGGTAAGGAGCAGCGTGGCTGGAAACGTTCTTCTGACATTTCA 346
714 BS.B.DNY...H..YNNY.HRNB.Y.RCA.N.NC..WSCMH.RA.YDD.SMNSBW..T.S 655
347 GAACAGGGGCTACAGTCTCGATATAGATATAGAGCGCATTTTGTCTAGCGCGCGG 406
654 SBDSYNCB.A...W.RSNN..M.TWMTS.HR..D..A...YN.TAANC..A.B.RCK 595
407 GCGCCGTTTCCCAATAGGAGCGCAGTTTATCGGCGAGCTCTACTTCTTCTATTG 466
594 ..AM.HKB...NBRMCNYM.T.S.ANW.HNYTTCRMD.A.RH..CY..BSDCK.NT.KY 535
467 GGTAAAGCCCTTCTGTCTTGGCGCAGTGTCTGACAGCTCGCGCGAGAACATAGTG 526
534 ..TTCMRT.WH.BYH.HAT..B..SB.DSST.CH..KB.D.BHA..Y.M.KWNC.R.RH 475
527 ATAAGGATGTAACTTTCATGATGAGATTAAGCAAGCGAAAAAATAGTGGCTAGCTG 586
474 ACG...RDT.C.M.WBH.WMB.AB.HC.W.DG..DB.BABH..RS.SBS.H..Y..B 415
587 GAGTTGTTTCAATATATAAAGGAGAAATGTTGCTCACTATGTAAGCTTTCTGG 646
414 ..S.CWN.KTTR.DM.BT..H.M..SR.BM.RH.R...YMA.D.TRHSCYV.TH.YTR 355
647 GACGTTTAACTTTTATGAGAGAGCATCAATCAATACAGATATGCAAAAAA 706
354 .YBBSCMAA.YRA.NG.TWT.RRASK..A.S.DHABWCH..BSA.B.KUS.S.RAT. 295
```

```
QY 707 AGACTAATATAAC 721
DB 294 DM.S.M.ATWR...H 280
```

## RESULT 13

US-10-140-923-198/c  
Sequence 198, Application US/10140923  
Publication No. US20030207355A1  
GENERAL INFORMATION:

```
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerltsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C18
CURRENT APPLICATION NUMBER: US/10/140,923
CURRENT FILING DATE: 2002-05-07
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 198
LENGTH: 1024
TYPE: PRT
ORGANISM: Homo Sapien
US-10-140-923-198
```

Query Match 5.7%; Score 41; DB 17; Length 1024;

Best Local Similarity 7.7%; Pred. No. 0.086; Matches 52; Conservative 194; Mismatches 429; Indels 0; Gaps 0;

```
47 TACACTACGAAAAACCATTAAGAGCAAGATCTACTTGGAGAAAGAGAGCAG 106
954 T.CRYAN..MMMH..MWT.HYCD.NYTD..BH.TH.H.BB.SN.S.N..YN..NYHS 895
107 CTGTGAAGGGAGTGGGGCTAAGAGTCACTTCTTTCCCTCGCGCGGAC 166
894 GA.T...MM.SBTAASD.MKW..WBH.H..SBQAGR.HYBN...YNYNS.WHS..A 835
167 CCGGAGCCCTCTCTCCCGCAGATTTCTCTTCAATATCTCTTATTCATC 226
834 ST..DGMNC.SGT.K.B.YY..D.RST.B.SCN.NNM.WC.M.RATB.M..R.N..N 775
227 CCGTTGAAGCAACCGCATATGACTAATGCTGACATCTCCATGCTGACTGT 286
774 RN.KYH..MH..R.TYSTTDW...HM..S.RY.....N..RCTYT.S.THH..CTYNS 715
287 GTGTATCTACAGTGGTAAGGAGCAGCGTGGCTGGAAACGTTCTTCTGACATTTCA 346
714 BS.B.DNY...H..YNNY.HRNB.Y.RCA.N.NC..WSCMH.RA.YDD.SMNSBW..T.S 655
347 GAACAGGGGCTACAGTCTCGATATAGATATAGAGCGCATTTTGTCTAGCGCGCGG 406
654 SBDSYNCB.A...W.RSNN..M.TWMTS.HR..D..A...YN.TAANC..A.B.RCK 595
407 GCGCCGTTTCCCAATAGGAGCGCAGTTTATCGGCGAGCTCTACTTCTTCTATTG 466
594 ..AM.HKB...NBRMCNYM.T.S.ANW.HNYTTCRMD.A.RH..CY..BSDCK.NT.KY 535
```

```

QY 467 GGTAAAGCCCTTCTGTTTCGCGCAGTGGTGTGAGGCTGCGCGGAGAACATAGTG 526
DB 534 ..TTGMMRT.WH.BYH.HAT..B..SB.DSST.CH..KB.D.BHA..Y.M.KWNC.R.RH 475
QY 527 ATAGAGGATTAACCTTTCGATGAGAGATTAAGCAGCGGAGAAAACATATGCTAGCTGG 586
DB 474 ACB...RDT.C.M.WBH.WWB.AB.HC.W.DG..DB.BKABH..RS.SSBS.H..Y..B 415
QY 587 GAGTTGTTTTCATCATATATAAGGAGAAATGTTGCTCACTATGTGACAGTTTCTGG 646
DB 414 ..S.CWN.KTTR.DM.BT..H.M..SR.BM.RH.R...YMA.D.TRHSCY.YH.YTR 355
QY 647 GAGCTTAACTTTATGAGAGAGATCAATCATATGATATGTCAAAAAAAA 706
DB 354 ..YBBSCMAAA.YRA.NG.TWT.RRASK..A.S.DHABWCH...BSA.B.KNS.S.RAT. 295
QY 707 AAGACTAATATAAC 721
DB 294 DM.S.M.ATWR...H 280

```

```

RESULT 14
US-10-141-756-198/c
; Sequence 198, Application US/10141756
; Publication No. US20030207359A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C200
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 198
; LENGTH: 1024
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-141-756-198

```

```

Query Match 5.7%; Score 41; DB 17; Length 1024;
Best Local Similarity 7.7%; Pred. No. 0.088;
Matches 52; Conservative 194; Mismatches 429; Indels 0; Gaps 0;
QY 47 TACACTACGAAACCATTAAGACGAAAGGATCTTCTTGAAGGAAAGAGACAG 106
DB 954 T..CRYAN..MMWY..MT..HYCD..NYTBD..BH..TH.H.BB..SN.S.N..YN..NYHS 895
QY 107 CTGTGAAGGGGATGAGGAGTCAATGATCTTCTTCCCTGCGGCGGAC 166
DB 894 GA.T...MM.SPTASD.MKW...WBH.H..SBCAGR.HYEN...YATNSG.WHS..A 835
QY 167 CCGGAGCCCTCTCTCCCGCAGCATTTCTTCTTCAATATCTCTTTTATCTATC 226
DB 834 ST..DSDMNC.SGT.K.B.YY...D..RST.B.SCN.NNM.WC.M.RAATB.M..R.N..N 775
QY 227 CCGTTAAGCAACGCACTATGACTAATAGTGTGACATCTCCATGCGTGTGACTTGT 286

```

```

DB 774 RN.KTH..MH..R.TYSTDW...HM..S.RY.....N...RCYTT.S.THH..CTYNS 715
QY 287 GTGTATCTCAAGATGTGAACGCGACCGTCTCGAAGACGTTCTTCTGTGACATTTCTA 346
DB 714 BS.B.DNY...H...YMWY.HNRBY.RCA.N.NC..WSCMH.RA.YDD.SPNSBW..T.S 655
QY 347 GAACAGGCGCTACAGTCTGATTAATGAAATATATAGCGCATTTTGTGCGCGCGCG 406
DB 654 SBDSYNCB.A...W.RENN..M.TMAMTS..HR..D..A...YN.TAANC..A.B.RCK 595
QY 407 GCGCCGTTTCCCAATAGGAGCGGAGCTTATGCGGAGAGCTCTTCTCTTTTG 466
DB 594 ..AW.HKB...NBRKCNM.T.S.ANW..HNYTNCMD.A.RH..CY..BSPDK..NT..KY 535
QY 467 GGTAAAGCCCTTCTGTTTCGCGCAGTGGTGTGAGGCTGCGCGGAGAACATAGTG 526
DB 534 ..TTGMMRT.WH.BYH.HAT..B..SB.DSST.CH..KB.D.BHA..Y.M.KWNC.R.RH 475
QY 527 ATAGAGATTAACCTTTCGATGAGAGATTAAGCAGCGGAGAAAACATATGCTAGCTGG 586
DB 474 ACB...RDT.C.M.WBH.WWB.AB.HC.W.DG..DB.BKABH..RS.SSBS.H..Y..B 415
QY 587 GAGTTGTTTTCATCATATATAAGGAGAAATGTTGCTCACTATGTGACAGTTTCTGG 646
DB 414 ..S.CWN.KTTR.DM.BT..H.M..SR.BM.RH.R...YMA.D.TRHSCY.YH.YTR 355
QY 647 GAGCTTAACTTTATGAGAGAGATCAATCATATGATATGTCAAAAAAAA 706
DB 354 ..YBBSCMAAA.YRA.NG.TWT.RRASK..A.S.DHABWCH...BSA.B.KNS.S.RAT. 295
QY 707 AAGACTAATATAAC 721
DB 294 DM.S.M.ATWR...H 280

```

```

RESULT 15
US-10-141-759-198/c
; Sequence 198, Application US/10141759
; Publication No. US20030207361A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C197
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 198
; LENGTH: 1024
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-141-759-198

```

```

Query Match 5.7%; Score 41; DB 17; Length 1024;
Best Local Similarity 7.7%; Pred. No. 0.088;

```

